

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, G01N 33/53, C12P 19/34, C12N 5/10, 1/21, C07K 5/00, 14/00, 16/00, C07H 21/02, 21/04		A1	(11) International Publication Number: WO 96/25519
			(43) International Publication Date: 22 August 1996 (22.08.96)
(21) International Application Number: PCT/US96/01938			(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AZ, BY, KG, KZ, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
(22) International Filing Date: 15 February 1996 (15.02.96)			
(30) Priority Data: 08/390,878 17 February 1995 (17.02.95) US			
(71) Applicant: PATHOGENESIS CORPORATION [US/US]; Suite 150, 201 Elliott Avenue West, Seattle, WA 98119 (US).			
(72) Inventors: STOVER, Charles, Kendall; 7640 81st Place S.E., Mercer Island, WA 98040 (US). MAHAIRAS, Gregory, G.; 3312 39th West, Seattle, WA 98199 (US).			
(74) Agents: HUNTER, Tom et al.; Townsend and Townsend and Crew, Stewart Street Tower, One Market, San Francisco, CA 94105-1492 (US).			

Published*With international search report.***(54) Title: VIRULENCE-ATTENUATING GENETIC DELETIONS****(57) Abstract**

The present invention provides specific genetic deletions that result in an avirulent phenotype of a mycobacterium. These deletions may be used as phenotypic markers of providing a means for distinguishing between disease-producing and non-disease producing mycobacteria.

IDS

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

VIRULENCE-ATTENUATING GENETIC DELETIONS

BACKGROUND OF THE INVENTION

Mycobacterium tuberculosis (MTB) infects over ten million people each year and kills over three million, making it the infectious agent causing the greatest mortality worldwide. In an effort to combat *Mycobacterium tuberculosis*, vaccination programs using a viable attenuated strain of *Mycobacterium bovis* called bacille Calmette-Guérin (BCG) have been established in more than 120 countries over the course of the last 5 decades. Although widely used and considered safe enough to administer to infants, the BCG vaccine is controversial for two principle reasons: 1) Efficacy for BCG vaccines against tuberculosis has varied from 0-85% in different clinical trials; and 2) Immunization with BCG sensitizes vaccinees to the tubercular antigens used in the tuberculin skin test, confounding attempts to discriminate between BCG immunization and TB infection. For these two reasons, especially the latter, BCG is not used in the United States where surveillance with the tuberculin test is preferred.

The original Pasteur BCG strain was developed by multiple (230 times) serial passages in liquid culture. BCG has never been shown to revert to virulence in animals indicating that the attenuating mutations in BCG are stable deletions and/or multiple mutations which cannot revert. However, the mutations which arose during serial passage of the original BCG strain have never been identified. Moreover, recent efforts to genetically complement BCG virulence with genomic libraries of virulent tubercle bacilli have also been unsuccessful again suggesting that multiple unlinked mutations are responsible for the attenuation of BCG virulence. The antigenicity of BCG and the characteristics leading to its avirulence are thus poorly understood.

SUMMARY OF THE INVENTION

The present invention provides specific genetic deletions that account for the avirulent phenotype of the bacille Calmette-Guérin (BCG) strain of *Mycobacterium bovis*. These deletions may be used as phenotypic markers of providing a means for distinguishing between disease-producing and non-disease producing mycobacteria.

In a preferred embodiment, this invention provides for nucleic acid sequences that are markers for avirulent or virulent mycobacteria. The sequences uniquely characterize the presence or absence of deletions that result in an avirulent phenotype. More specifically the sequence are either deletion junction sequence or deletion sequences or subsequences within deletion junction sequences or deletion sequences. Thus, this invention provides for a marker for an avirulent mycobacterium comprising a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement includes BCG Δ 1a, BCG Δ 1b, BCG Δ 2a, BCG Δ 2b, BCG Δ 3a, BCG Δ 3b, BCG Δ 1ab, BCG Δ 2ab, BCG Δ 3ab, BCG Δ 1, BCG Δ 2, and BCG Δ 3. In a particularly preferred embodiment, the marker specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from *Mycobacterium tuberculosis* or *Mycobacterium bovis*, or alternatively, the marker specifically hybridizes under stringent conditions to a nucleic acid from *Mycobacterium tuberculosis* or *Mycobacterium bovis*, but not to a nucleic acid from BCG. The marker may be the full length BCG Δ 1a, BCG Δ 1b, BCG Δ 2a, BCG Δ 2b, BCG Δ 3a, BCG Δ 3b, BCG Δ 1ab, BCG Δ 2ab, BCG Δ 3ab, BCG Δ 1, BCG Δ 2, and BCG Δ 3 or a subsequence within any of these regions. The marker may also include a nucleic acid having at least 80%, preferably 90%, more preferably 95%, and most preferably 98% percent sequence identity with BCG Δ 1a, BCG Δ 1b, BCG Δ 2a, BCG Δ 2b, BCG Δ 3a, BCG Δ 3b, BCG Δ 1ab, BCG Δ 2ab, BCG Δ 3ab, BCG Δ 1, BCG Δ 2, or BCG Δ 3. The marker may also include a sequence selected from an open reading frame of a the deletion sequences BCG Δ 1, BCG Δ 2, BCG Δ 3. Suitable open reading frames are indicated in Figures 4, 5, and 6.

The above described marker may be a probe. The probe may be labeled by a number of means including, but not limited to radioactive, fluorescent, enzymatic, and colorimetric labels.

In another embodiment, this invention provides for polypeptides encoded by a subsequence of the BCG Δ 1, BCG Δ 2, or BCG Δ 3 deletions. In particular, the subsequence may be selected from an open reading frame (ORF) present in one of these deletion sequences. This invention also provides for monoclonal or polyclonal antibodies that

specifically bind polypeptides encoded by one or more subsequences of the BCG Δ 1, BCG Δ 2, or BCG Δ 3 deletions.

In still another embodiment, this invention provides for a recombinant cell comprising a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement is BCG Δ 1a, BCG Δ 1b, BCG Δ 2a, BCG Δ 2b, BCG Δ 3a, BCG Δ 3b, BCG Δ 1ab, BCG Δ 2ab, BCG Δ 3ab, BCG Δ 1, BCG Δ 2, or BCG Δ 3. The recombinant cell may be a mycobacterium. The recombinant cell may express a polypeptide encoded by any of BCG Δ 1a, BCG Δ 1b, BCG Δ 2a, BCG Δ 2b, BCG Δ 3a, BCG Δ 3b, BCG Δ 1ab, BCG Δ 2ab, BCG Δ 3ab, BCG Δ 1, BCG Δ 2, and BCG Δ 3. More preferably, the recombinant cell expresses a polypeptide encoded by an intact open reading frame present in any of these regions. The cell may also be a mycobacterium having one or more deletions in the BCG Δ 1, BCG Δ 2, or BCG Δ 3 genomic regions where the deletions result in the attenuation of an otherwise virulent strain of mycobacterium and wherein the deletions are present in up to two of the genomic regions.

In still yet another embodiment, this invention provides a method of distinguishing between an attenuated and a virulent mycobacterium. The method involves detecting the presence or absence of a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement is BCG Δ 1a, BCG Δ 1b, BCG Δ 2a, BCG Δ 2b, BCG Δ 3a, BCG Δ 3b, BCG Δ 1ab, BCG Δ 2ab, BCG Δ 3ab, BCG Δ 1, BCG Δ 2, or BCG Δ 3. The first nucleic acid may include any of the markers described above. A particularly preferred marker is one that specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from *Mycobacterium tuberculosis* or *Mycobacterium bovis*, or alternatively, that specifically hybridizes under stringent conditions to a nucleic acid from *Mycobacterium tuberculosis* or *Mycobacterium bovis*, but not to a nucleic acid from BCG. The method may involve amplifying either the first nucleic acid by any of a number of methods including, for example, polymerase chain reaction. The detection may involve detecting the first nucleic acid, for example, as in a Southern blot, or alternatively, detecting a polypeptide encoded by the first nucleic acid. More specifically, the polypeptide may be

a encoded by an open reading frame (ORF) selected from BCG Δ 1, BCG Δ 2, or BCG Δ 3. The polypeptide may be visualized by a number of means well known to those of skill in the art including antibody hybridization such as direct or indirect binding of labeled antibody.

5 This invention additionally provides a method for determining whether an attenuated or a virulent *Mycobacterium* is present in a sample. This method involves providing a first nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a complement of the second nucleic acid where the second nucleic acid or its complement is BCG Δ 1a, BCG Δ 1b, BCG Δ 2a, BCG Δ 2b, BCG Δ 3a, BCG Δ 3b, BCG Δ 1ab, 10 BCG Δ 2ab, BCG Δ 3ab, BCG Δ 1, BCG Δ 2, or BCG Δ 3; and hybridizing the first nucleic acid to the biological sample. The first nucleic acid may include any of the markers described above. A particularly preferred marker is one that specifically hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from *Mycobacterium tuberculosis* or *Mycobacterium bovis*, or alternatively, that specifically hybridizes under 15 stringent conditions to a nucleic acid from *Mycobacterium tuberculosis* or *Mycobacterium bovis*, but not to a nucleic acid from BCG. The method may involve amplifying either the first nucleic acid by any of a number of methods including, for example, polymerase chain reaction. The detection may involve detecting the first nucleic acid, for example, as in a Southern blot, or alternatively, detecting a polypeptide encoded by the first nucleic acid. 20 More specifically, the polypeptide may be a encoded by an open reading frame (ORF) selected from BCG Δ 1, BCG Δ 2, or BCG Δ 3. The method may also include detecting the hybridized first nucleic acid. This may involve direct detection of a label or additionally involve an amplification step and subsequent detection of the amplified product.

Finally, this invention provides a method of producing an attenuated-virulence 25 mycobacterium. This method involves deleting from the genomic DNA of a virulent mycobacterium a first nucleic acid that specifically hybridizes under stringent conditions with a second nucleic acid or a complement of said second nucleic acid where said second nucleic acid or complement of said second nucleic acid is selected from the group consisting of BCG Δ 1, BCG Δ 2, and BCG Δ 3. The first nucleic acid may be BCG Δ 1, BCG Δ 2, or BCG Δ 3, 30 or alternatively, it may be a promoter, other control element or an open reading frame from BCG Δ 1, BCG Δ 2, or BCG Δ 3.

Definitions

Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. For purposes of the present invention, the following terms are defined below.

The phrase "specifically detect" as used herein refers to the process of determining that a particular subsequence is present in a DNA sample. A DNA sequence may be specifically detected through a number of means known to those of skill in the art. These would include, but are not limited to amplification of the particular target sequence through polymerase chain reaction or ligase chain reaction, hybridization of the sequence to a labeled probe, and binding by labelled ligands or monoclonal antibodies. For a discussion of various means of detection of specific nucleic acid sequences see Perbal, B. *A Practical Guide to Molecular Cloning*, 2nd Ed. John Wiley & Sons, N.Y. (1988) which is incorporated herein by reference.

The phrase "select subsequence" is used herein to refer to a particular DNA subsequence that is of interest. It is often a predetermined or known sequence of nucleic acid bases. A select subsequence is typically chosen because of a unique sequence identity. Typically a select subsequence is targeted for DNA amplification and often is useful as a specific marker for the presence of a particular gene or a deletion of a particular nucleic acid sequence.

The term "oligonucleotide" refers to a molecule comprised of two or more deoxyribonucleotides or ribonucleotides. Oligonucleotides may include, but are not limited to, primers, probes, nucleic acid fragments to be detected, and nucleic acid controls. Oligonucleotides include naturally occurring nucleotides, chemically modified naturally occurring nucleotides and synthetic nucleotides. The exact size of an oligonucleotide depends on many factors and the ultimate function or use of the oligonucleotide.

The term "primer" refers to an oligonucleotide, whether natural or synthetic, capable of acting as a point of initiation of DNA synthesis under conditions in which synthesis of a primer extension product complementary to a nucleic acid strand is induced, *i.e.*, in the presence of four different nucleoside triphosphates and an agent for polymerization (*i.e.*, DNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. A primer is preferably a single-stranded oligodeoxyribonucleotide.

The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 25 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to
5 hybridize with a template.

The phrase "PCR primers competent to amplify" as used herein refers to a pair of PCR primers whose sequences are complementary to DNA subsequences immediately flanking the DNA subsequence (target sequence) which it is desired to amplify. The primers are chosen to bind specifically those particular flanking subsequences and no other sequences
10 present in the sample. The PCR primers are thus preferably chosen to amplify the unique target sequence and no other. Alternatively, the PCR primers may be selected to bind to sequences other than the target sequence where the amplification products can be subsequently distinguished (*e.g.* where the desired amplified sequence is different in size than other amplified sequences).

"Amplifying" or "amplification", which typically refer to an "exponential" increase in target nucleic acid, are used herein to describe both linear and exponential increases in the number of a select target sequence of nucleic acid.
15

The term "antisense orientation" refers to the orientation of nucleic acid sequence from a structural gene that is inserted in an expression cassette in an inverted manner with respect to its naturally occurring orientation. When the sequence is double
20 stranded, the strand that is the template strand in the naturally occurring orientation becomes the coding strand, and vice versa.

The term "deletion" refers to a region of a nucleic acid which is not present in an organism, but which is present in another related organism. In the context of mycobacteria, a deletion refers, *e.g.*, to a region of nucleic acid which is not present in one
25 strain of mycobacteria, but which is present in another related strain. For instance, an avirulent mycobacterial strain can have a deletion in its genome relative to the genome of a related virulent mycobacterial strain.

The term "deletion junction" refers to the region of a nucleic acid spanning the insertion point of a deletion. Thus, where a region of a nucleic acid sequence is deleted (*i.e.* a deletion is present), the deletion junction spans the nucleotides that are immediately
30 adjacent to the deletion. Conversely, where a region of a nucleic acid sequence is not

deleted (*i.e.* the deletion is absent), two deletion junctions are present, each spanning respectively one end of the deletion sequence and its flanking sequence.

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or gene sequence given in a sequence listing, such as a polynucleotide sequence of Figures 1, 2, or 3, or may comprise a complete cDNA or gene sequence.

Generally, a reference sequence is at least 10 nucleotides in length, frequently at least 20 to 25 nucleotides in length, and often at least 50 nucleotides in length. Sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window", as used herein, refers to a segment of at least 10 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 10 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences.

Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman *Adv. Appl. Math.* 2: 482 (1981), by the homology alignment algorithm of Needleman and Wunsch *J. Mol. Biol.* 48: 443 (1970); by the search for similarity method of Pearson and Lipman *Proc. Natl. Acad. Sci. (USA)* 85: 2444 (1988), or by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection, and the best alignment (*i.e.*, resulting in the highest percentage of sequence similarity over the comparison window) generated by the various methods is selected.

The term "sequence identity" means that two polynucleotide sequences are identical (*i.e.*, on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned

sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "identical" in the context of two nucleic acid or polypeptide sequences refers to the residues in the two sequences which are the same when aligned for maximum correspondence.

The terms "isolated" or "biologically pure" refer to material which is substantially or essentially free from components which normally accompany it as found in its native state. The isolated nucleic acid probes of this invention do not contain materials normally associated with their *in situ* environment, in particular nuclear, cytosolic or membrane associated proteins or nucleic acids other than those nucleic acids intended to comprise the nucleic acid probe itself.

The term "marker" refers to a characteristic which distinguishes one class of cells or compositions from a second class of cells or compositions. For instance, the deletions and deletion junctions described herein can be used to distinguish between strains (e.g., virulent and avirulent strains) of mycobacteria. While markers are indicators of associated features or properties, as used herein, markers may also be used for purposes other than indicating the associated feature or property. Thus, for example, a nucleic acid marker of virulence identifies a particular nucleic acid which may be used in a variety of contexts other than simply indicating virulence.

The term "nucleic acid" refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompassing known analogues of natural nucleotides that can function in a similar manner as naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence includes the complementary sequence thereof.

The term "operably linked" refers to functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates transcription of RNA corresponding to the second sequence.

The term "peptide" or "polypeptide" refers to an amino acid polymer which is encoded by a nucleic acid. The peptide or polypeptide may include naturally occurring or modified amino acids.

The terms "probe" or "nucleic acid probe" refer to a molecule that binds to a specific sequence or subsequence of a nucleic acid. A probe is preferably a nucleic acid which binds through complementary base pairing to the full sequence or to a subsequence of a target nucleic acid. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions. The probes are preferably directly labelled as with isotopes, chromophores, lumiphores, chromogens, or indirectly labelled such with, *e.g.*, biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the selected sequence or subsequence.

The term "labeled nucleic acid probe" refers to a nucleic acid probe that is bound, either covalently, through a linker, or through ionic, van der Waals or hydrogen "bonds" to a label such that the presence of the probe may be detected by detecting the presence of the label bound to the probe.

The term "recombinant" when used with reference to a cell indicates that the cell replicates or expresses a nucleic acid, or expresses a peptide or protein encoded by DNA whose origin is exogenous to the cell. Recombinant cells can express genes that are not found within the native (non-recombinant) form of the cell. Recombinant cells can also express genes found in the native form of the cell wherein the genes are re-introduced into the cell by artificial means.

The term "sample" refers to a material with which bacteria may be associated. Frequently the sample will be a "clinical sample" which is a sample derived from a patient. Such samples include, but are not limited to, sputum, blood, blood cells (*e.g.*, white cells), tissue or fine needle biopsy samples, urine, peritoneal fluid, and pleural fluid, or cells therefrom. It will be recognized that the term "sample" also includes supernatant from eukaryotic cell cultures (which may contain free bacteria), cells from cell or tissue culture, and other media in which it may be desirable to detect mycobacteria (*e.g.*, food and water).

The term "subsequence" in the context of a particular nucleic acid sequence refers to a region of the nucleic acid equal to or smaller than the specified nucleic acid.

The term "substantial identity" or "substantial similarity" indicates that a nucleic acid or polypeptide comprises a sequence that has at least 90% sequence identity to a reference sequence, or preferably 95%, or more preferably 98% sequence identity to the

reference sequence, over a comparison window of at least about 10 to about 100 nucleotides or amino acid residues. An indication that two polypeptide sequences are substantially identical is that one protein is immunologically reactive with antibodies raised against the second protein. An indication that two nucleic acid sequences are substantially identical is that the polypeptides which the first nucleic acids encodes is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. Stringent conditions are sequence-dependent and will be different with different environmental parameters. Generally, stringent conditions are selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is at least about 0.2 molar at pH 7 and the temperature is at least about 60°C.

The term "uninterrupted reading frame" or "open reading frame" refers to a DNA sequence (*e.g.*, cDNA) lacking a stop codon or other intervening, untranslated sequence. An intact open reading frame refers to a full length uninterrupted reading frame or minor variations thereof.

The term "virulent" in the context of mycobacteria refers to a bacterium or strain of bacteria that replicates within a host cell or animal at a rate that is detrimental to the cell or animal within its host range. More particularly virulent mycobacteria persist longer in a host than avirulent mycobacteria. Virulent mycobacteria are typically disease producing and infection leads to various disease states including fulminant disease in the lung, disseminated systemic milliary tuberculosis, tuberculosis meningitis, and tuberculosis abscesses of various tissues. Infection by virulent mycobacteria often results in death of the host organism. Typically, infection of guinea pigs is used as an assay for mycobacterial virulence. In contrast, the term "avirulent" refers to a bacterium or strain of bacteria that either does not replicate within a host cell or animal within its host range, or replicates at a rate that is not significantly detrimental to the cell or animal.

The term BCG-like avirulence, as used herein refers to an attenuated virulence brought about by one of the deletions of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the complete sequence listing of the BCG deletion region 1 including flanking sequences. The deletion, designated BCG Δ 1, is located between nucleotide 2327 and nucleotide 11126.

5 Figure 2 shows the complete sequence listing of the BCG deletion region 2 including flanking sequences. The deletion, designated BCG Δ 2, is located between nucleotide 3382 and nucleotide 14071.

Figure 3 shows the complete sequence listing of the BCG deletion region 3 including flanking sequences. The deletion, designated BCG Δ 3, is located between
10 nucleotide 1406 and nucleotide 10673. "N" represents "A", "C", "G", or "T".

Figure 4 shows a map of the deletion sequence BCG Δ 1. This map identifies the various open reading frames (ORFs) and indicates their location within the deletion sequence. Ribosome binding sites and homologies to the predicted encoded proteins are shown.

15 Figure 5 shows a map of the deletion sequence BCG Δ 2. This map identifies the various open reading frames (ORFs) and indicates their location within the deletion sequence. Ribosomal binding sites and homologies to the predicted encoded proteins are shown.

Figure 6 shows a map of the deletion sequence BCG Δ 3. This map identifies
20 the various open reading frames (ORFs) and indicates their location within the deletion sequence. Ribosome binding sites and homologies to the predicted encoded proteins are shown. The sequence of a small region, estimated to be much less than 200 bp and located close to 9400 bp in Figure 3, remains to be determined. Therefore, the base pair coordinates given in the region 3 map 3' to the 9kb marker are approximations. The precise
25 sequence determination of this region is likely to effect the length of open reading frames 3H and 3L.

Figure 7 illustrates the deletion junction regions of BCG Δ 1, BCG Δ 2, and BCG Δ 3. The "terminal" deletion junction regions formed by the flanking sequences and the terminal regions of the deletion sequences are identified as BCG Δ 1a, BCG Δ 1b, BCG Δ 2a, BCG Δ 2b, and BCG Δ 3a, and BCG Δ 3b. When the deletion is present (the deletion sequences
30

are missing) the respective "a" and "b" sequences will be juxtaposed, thereby forming deletion "spanning" junction sequences designated BCG Δ 1ab, BCG Δ 2ab, and BCG Δ 3ab, respectively.

Figure 8 shows EcoRI and BamHI restricted chromosomal DNAs from *Mycobacterium bovis*, BCG Connaught, and *Mycobacterium tuberculosis* strains H37Ra, H37Rv, and Erdman probed with ³²P labeled BCG subtracted probe.

DETAILED DESCRIPTION

This invention reflects the discovery of genetic deletions in mycobacteria that result in an avirulent genotype such as is exhibited by the bacille Calmette-Guérin (BCG) mycobacterium. The original Pasteur bacille Calmette-Guérin (BCG) strain was developed by multiple (230 times) serial passages in liquid culture. BCG has never been shown to revert to virulence in animals indicating that the attenuating mutations in BCG are stable deletions and/or multiple mutations that cannot revert. The mutations that arose during serial passage of the original BCG strain were not previously known. Recent efforts to genetically complement BCG virulence with genomic libraries of virulent tubercle bacilli were unsuccessful, again suggesting that multiple unlinked mutations are responsible for the attenuation of BCG virulence.

The genetic deletions leading to the avirulent phenotype of BCG were identified by genomic subtractions between Connaught strain of BCG and MBV/MTB. The subtracted probe resulting from the genomic subtraction between BCG and the H37 Rv strain of *M. tuberculosis* was subsequently used to identify and clone three regions from a cosmid library of *Mycobacterium bovis* genomic DNA. Southern blot mapping and DNA sequence comparisons between BCG and *M. bovis* showed that three regions, designated regions 1-3, contained DNA segments of approximately 9 kb, 11 kb and 9 kb respectively, which are deleted in the Connaught strain of BCG. Precise deletion junctions were identified for each region by comparisons of BCG and corresponding virulent MBV sequences. The respective deletions, designated BCG Δ 1, BCG Δ 2 and BCG Δ 3 are illustrated in Figures 1-3.

One of skill in the art will appreciate that the deletions encompassed by BCG Δ 1, BCG Δ 2 and BCG Δ 3 may be utilized in a variety of contexts. For example, the deletions may be utilized to distinguish between avirulent and virulent strains of

mycobacteria thereby providing early detection of patients at risk for tuberculosis. This is of particular importance where mycobacteria are identified in a sample from a patient that has been previously vaccinated with BCG. In this context it may be critical to determine whether mycobacteria identified in a biological sample from such a patient are pathogenic.

5 In another embodiment, the preparation of mycobacteria containing the deletions of the present invention may provide superior vaccines to BCG which has long been known to have marginal efficacy. Thus, for example, a *Mycobacterium tuberculosis* may contain a full BCGΔ1 deletion or a smaller deletion within BCGΔ1 (e.g. one or more open reading frames) rendering it avirulent. An avirulent MTB will provide a more efficient
10 vaccine because it is antigenically more similar to MTB than is BCG. Moreover, an MTB rendered avirulent by the production of smaller deletions within the deletion regions identified in this invention will present more antigenic determinants.

Since the loss of virulence is due to the loss of gene products expressed by the nucleic acid sequences comprising the deletion regions, the BCGΔ1, BCGΔ2 and BCGΔ3
15 deletion sequences and proteins encoded within these deletion sequences provide suitable targets for drug screening. Thus, the use of deleted sequences as targets to screen for drugs that inhibit or interfere with transcription, translation, or post-translational processing of proteins encoded by the deletion sequences, or with the deletion encoded polypeptides themselves, provides an assay for anti-mycobacterial agents. In particular, the use of
20 reporter genes such as firefly luciferase (FFlux), β-galactosidase (BGal), and the like, under the control of promoters present in the deletion sequence provide a rapid assay for drugs regulating activity originating in this region. Conversely, since the protein products of the deletion sequences are presumably expressed in virulent mycobacterial species, proteins expressed by deletion sequences may make good antigens for antimycobacterial vaccines.

25 Finally, as the viability of BCG demonstrates, deletion regions BCGΔ1, BCGΔ2 and BCGΔ3 are not required for mycobacterial growth and reproduction. Thus, these deletion regions provide good insertion points for the expression of heterologous DNA. The heterologous DNA sequences may be under the control of endogenous inducible or constitutive promoters typically found in the deletion sequences, or alternatively, they may
30 be under the control of introduced promoters, either constitutive or inducible, exogenous to mycobacteria.

I. Detection of Deletions

As indicated above, the deletions identified in the present invention provide useful markers for the identification of an avirulent (or conversely a virulent) mycobacterial phenotype. Specifically, determination of avirulence simply requires the detection of the presence or absence of the deletion (either BCG Δ 1, BCG Δ 2, or BCG Δ 3, or deletions within these regions). Where the deletion is present in the bacterial DNA, the bacterium expresses a BCG-like avirulent phenotype. Conversely, where the deletion is absent in the bacterial DNA, the bacterium does not express a BCG-like avirulence. While this may indicate that the bacterium is virulent, one of skill will appreciate that the bacterium may still be avirulent due to the presence of other mutations or deletions. Nevertheless, screening for the presence of the deletion provides a means of detecting a BCG-like avirulent mycobacterium.

Means of detecting deletions are well known to those of skill in the art. Generally, the deletions may be detected either by detecting the presence or absence of deletion junctions, or, alternatively, by detecting the presence or absence of the sequences contained within the deletion (deletion sequences). Where a nucleic acid sequence is deleted (*i.e.*, a deletion is present), the sequences that previously flanked the deleted sequence are juxtaposed, thereby forming a new deletion junction that spans the deletion. Detection of the presence of such a "spanning" deletion junction indicates the presence of the deletion and thus the avirulent phenotype.

Conversely, where the nucleic acid sequence is not deleted (the deletion is not present) the spanning junction sequence will be absent (See, *e.g.* Figure 7). The "terminal" deletion junction sequences flanking each endpoint of the deletion region are present and detection of these terminal deletion junctions indicates the absence of a deletion. Spanning deletion junction regions and terminal deletion junctions suitable for detecting the deletions of the present invention are illustrated in Figure 7 and in Table 1.

Table 1. Nucleic acid sequences comprising deletion junctions. The symbol "|" indicates the insertion point of the deletion sequence. Deletion sequence bases are represented in lower case letters.

Junction	Nucleotide Sequence	Seq. ID
BCG Δ 1a	CTGGTCGACGATTGGCACAT gcagccgtgggtgccgccgg	1

BCG Δ 1b	gtgtcttcacgggtccac CCAGCCGCCCGGATCCAGCA	2
BCG Δ 2a	CAACTCCACGGCGACCACCC gcgccccgctcgactaga	3
BCG Δ 2b	gcccacccggtcgagcaccc CGATGATCTTCTGTTTGACC	4
BCG Δ 3a	CACCTCGACCACGGCCAACC gtggacctgtgatacact	5
BCG Δ 3b	tcagcagtcacggccaacc CCGCACCAACACCTTCCACC	6
BCG Δ 1ab	CTGGTCGACGATTGGCACAT CCAGCCGCCCGGATCCAGCA	7
BCG Δ 2ab	CAACTCCACGGCGACCACCC CGATGATCTTCTGTTTGACC	8
BCG Δ 3ab	CACCTCGACCACGGCCAACC CCGCACCAACACCTTCCACC	9

Where a deletion is detected by determining the presence or absence of sequences contained within the deletion (deletion sequences), the absence of deletion sequences indicates the presence of a deletion and thus an avirulent phenotype. Conversely, the presence of deletion sequences indicates the absence of a deletion. Deletion sequences that provide suitable targets for detecting the deletions of the present invention are provided in Figures 1, 2 and 3.

A) Isolation of DNA for Detection of Mycobacterium Genomic Deletions

In a preferred embodiment, DNA is obtained from mycobacteria. As used herein, the term "mycobacteria" refers to any bacteria of the family *Mycobacteriaceae* (order *Actinomycetales*) and includes, but is not limited to, *Mycobacterium tuberculosis*, *Mycobacterium avium complex*, *Mycobacterium kansasii*, *Mycobacterium scrofulaceum*, *Mycobacterium bovis* and *Mycobacterium leprae*. These species and groups and others are described in Baron, S., ed. *Medical Microbiology*, 3rd Ed. (1991) Churchill Livingstone, New York, which is incorporated herein by reference.

The identification of deletions using a DNA marker requires that the DNA sequence be accessible to the particular probes used or to the components of the amplification system if the DNA sequence is to be amplified. In general, this accessibility is ensured by isolating the nucleic acids from the sample.

A variety of techniques for extracting nucleic acids from biological samples are known in the art. For example, see those described by Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New

York, (1985), by Han, *et al. Biochemistry*, 26: 1617-1625 (1987) and by Du, *et al. Bio/Technology*, 10: 176-181 (1992), which are incorporated herein by reference.

Alternatively, if the sample is readily disruptable, the nucleic acid need not be purified prior to amplification by the PCR technique, *i.e.*, if the sample is comprised of
5 cells, particularly peripheral blood lymphocytes or monocytes, lysis and dispersion of the intracellular components may be accomplished merely by suspending the cells in hypotonic buffer or boiling them in a low concentration of alkali (*i.e.* 10 mM NaOH).

In a preferred embodiment, DNA is extracted from mycobacteria as described in Example 1.

10

B) Detection of Deletions Using Hybridization Probes

In one embodiment the avirulence deletions are detected by contacting DNA obtained from the mycobacterium with a probe that specifically binds an entire deletion junction region or a subsequence of that region and does not specifically bind to any other
15 DNA sequences in the sample. Alternatively, a probe that specifically binds the entire deleted region or subsequence of that region and does not specifically bind to any other sequences in the sample is also suitable. While such probes may be proteins, oligonucleotide probes are preferred. Typically, the sequence of the oligonucleotide probe is chosen to be complementary to a select subsequence unique to the deletion junction or the
20 deletion sequence, whose presence or absence is to be detected. Under stringent conditions the probe will hybridize with the select subsequence forming a stable duplex.

The probe is typically labeled. Detection of the label in association with the target DNA indicates either the presence or absence of the deletion. The probe may be used to detect the deletion junction or deletion sequences directly in a DNA sample without
25 amplification of the deletion subsequences. In one embodiment, unamplified DNA sequences are probed using a Southern blot. The DNA of the sample is immobilized, on a solid substrate, typically a nitrocellulose filter or a nylon membrane. The substrate-bound DNA is then hybridized with the labeled probe under stringent conditions and non-specifically hybridized probe is washed away. Labeled probe detected in association with
30 the immobilized mycobacterial sequences (*e.g.* bound to the substrate) indicates the presence of deletion sequences (*e.g.* BCG Δ 1, BCG Δ 2, or BCG Δ 3) and therefore the absence of the deletion. Means for detecting specific DNA sequences are well known to those of skill in

the art. Protocols for Southern blots as well as other detection methods are provided in Maniatis, *et al. Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY (1982), which is incorporated herein by reference.

5 In another embodiment, the mycobacterial DNA subsequences are themselves labeled. They are then hybridized, under stringent conditions, with a probe immobilized on a solid substrate. Detection of the label in association with the immobilized probe indicates the presence or absence of the deletion.

10 In a preferred embodiment, the deletion junction sequences or subsequences or the deletion sequences or subsequences may be amplified by a variety of DNA amplification techniques (for example via cloning, polymerase chain reaction, ligase chain reaction, transcription amplification, *etc.*) prior to detection using a probe. Because the copy number of mycobacterial sequences bearing the virulence-attenuating deletions is low, the use of unamplified mycobacterial DNA results in an assay of low sensitivity. Amplification of mycobacterial DNA increases sensitivity of the assay by providing more
15 copies of possible target subsequences. In addition, by using labeled primers in the amplification process, the mycobacterial DNA sequences are labeled as they are amplified.

C) Selection of Probes for Detection of the Deletion Junction Sequences or the Deletion Sequences

20 Full length sequences are provided for the deletions BCG Δ 1, BCG Δ 2, and BCG Δ 3 in Figures 1, 2 and 3 respectively. Using these sequence listings, one of skill in the art may easily determine appropriate probes or primers for the detection of the presence or absence of the deletion junctions or the deletion sequences. Generally speaking, a probe will be selected that hybridizes to the target junction sequences or deletion sequences, but
25 not to other mycobacterial nucleic acid sequences under stringent conditions. The design of hybridization probes is well known in the art. See, for example, Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1989), which is incorporated herein by reference.

30 In a preferred embodiment, the probe is an oligonucleotide sequence complementary to a subsequence comprising a deletion junction (*e.g.* BCG Δ 1a, BCG Δ 1b, BCG Δ 2a, BCG Δ 2b, BCG Δ 3a, BCG Δ 3b, BCG Δ 1ab, BCG Δ 2ab, and BCG Δ 3ab) or a

sequence complementary to a subsequence of a deletion sequence (e.g. BCG Δ 1, BCG Δ 2, and BCG Δ 3). The probe preferably has destabilizing mismatches with subsequences from other regions of the mycobacterial genome.

5 The exact length of the probe depends on many factors including the length of conserved regions around the deletions, the degree of sequence specificity desired, and the amount of internal complementarity within the probe. Such probes are preferably 17 to 25 bases in length. One of skill will recognize that longer probes specifically hybridize at higher temperatures. Generally, stringent conditions are selected to be about 5°C to 20°C, more preferably about 10°C, lower than the thermal melting point (T_m) for the specific
10 sequence at a defined ionic strength and pH. Under stringent conditions, the probe will specifically hybridize to a nucleic acid sequence from an avirulent mycobacterium such as BCG, but not to a nucleic acid sequence from a virulent mycobacterium such as MTB or MBV. Alternatively, Under stringent conditions, the probe will specifically hybridize to a nucleic acid sequence from a virulent mycobacterium such as MTB or MBV, but not to a
15 nucleic acid sequence from an avirulent mycobacterium such as BCG.

Oligonucleotide probes can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphotriester method of Narang *et al.* *Meth. Enzymol.*, 68: 90-99 (1979); the phosphodiester method of Brown *et al.*, *Meth. Enzymol.*
20 68:109-151 (1979); the diethylphosphoramidite method of Beaucage *et al.*, *Tetra. Lett.*, 22: 1859-1862 (1981); and the solid support method of U.S. Patent No. 4,458,066.

Probe detectability may be increased by the attachment of a label. As used herein, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in
25 the present invention include magnetic beads (e.g. DynabeadsTM), fluorescent dyes (e.g., fluorescein isothiocyanate, texas red, rhodamine, and the like), radiolabels (e.g., ³H, ¹²⁵I, ³⁵S, ¹⁴C, or ³²P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g. polystyrene, polypropylene, latex, etc.) beads.

30 Methods for attaching labels to probes, primers, and antibodies are well known to those of skill in the art. For example, the probe can be labeled at the 5'-end with ³²P by incubating the probe with ³²P-ATP and polynucleotide kinase (see Perbal, A

Practical Guide to Molecular Cloning, 2nd ed. John Wiley, N.Y. (1988)). Other labels may be joined to the probe directly or through linkers. They may be located at the ends of the probe or internally. Methods of attaching labels may be found in Connell, *et al.*, *Bio/Techniques* 5: 342 (1987), U.S. Patent Nos. 4,914,210, 4,391,904 and 4,962,029, which are incorporated herein by reference. In addition, kits for labelling oligonucleotides are widely available. See, for example, Boehringer Mannheim Biochemicals (Indianapolis, IN) for "Genius" labeling kits based on dioxigenin technology and Clontech (South San Francisco, CA) for a variety of direct and indirect oligonucleotide labeling reagents.

D) Detection of Deletions Conferring Avirulence Through Amplification of Unique Subsequences

Deletions are particularly amenable to detection without the use of a hybridization probe. In a preferred embodiment, subsequences are amplified that include a deletion junction. The amplified deletion junction may be a "spanning" deletion junction in which case where the deletion is present (*i.e.* the deletion sequences are absent), the amplification product is a specific DNA incorporating the deletion junction sequence spanning the deletion (*e.g.* incorporating flanking sequences from both sides of the deleted sequence). Where the deletion is absent (*i.e.* deletion sequences are present) and primers are selected so that there are no priming sites within the deletion sequences, amplification is non-existent or alternatively provides a complex mixture of non-specifically amplified fragments. Alternatively, amplification primers may be selected that specifically hybridize to deletion sequences, as long as they are selected to amplify sequences that are distinguishable from the sequence amplified when the deletion is present.

Alternatively, the amplification product may be subsequence of a "terminal" deletion junction in which case absence of the deletion (*i.e.* the deletion sequences are present) will result in the amplification of the specifically targeted nucleic acid. Conversely, where the deletions are present (*i.e.* the deletion sequences are absent) there will be no specific amplification of a terminal deletion junction.

Amplification products may be separated by size for characterization. Size separation may be accomplished by a variety of means known to those of skill in the art.

These methods include, but are not limited to electrophoresis, density gradient centrifugation, liquid chromatography, and capillary electrophoresis. In a preferred embodiment, the fragments are separated by agarose gel electrophoresis. The bands are then stained with a marker to visualize them such as ethidium bromide and the gel is visualized, *e.g.*, using ultraviolet light.

As described above, an agarose gel typically shows 1 band if the deletion is present, reflecting amplification of the deletion-spanning sequence. Where the deletion is absent, amplification results in either no bands, where there are no sequences within the deletion to which the amplification primers may hybridize, or a smear where there is non-specific amplification, or a series of discrete bands distinguishable from the band representing the deletion-spanning sequence where primers are chosen that hybridize to deletion sequences.

E) Selection of Primers for Amplification of Avirulence Deletions

Amplification of deletion junction sequences or subsequences or deletion sequences or subsequences may be accomplished by methods well known in the art, which include, but are not limited to polymerase chain reaction (PCR) (Innis, *et al.*, *PCR Protocols. A guide to Methods and Application*. Academic Press, Inc. San Diego, (1990), which is incorporated herein by reference), ligase chain reaction (LCR) (see Wu and Wallace, *Genomics*, 4: 560 (1989), Landegren, *et al.*, *Science*, 241: 1077 (1988) and Barringer, *et al.*, *Gene*, 89: 117 (1990), which are incorporated herein by reference), transcription amplification (see Kwoh, *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 86: 1173 (1989) which is incorporated herein by reference), and self-sustained sequence replication (see Guatelli, *et al.*, *Proc. Nat. Acad. Sci. (U.S.A.)*, 87: 1874 (1990) which is incorporated herein by reference), each of which provides sufficient amplification so that the target sequence can be detected by nucleic acid hybridization to a probe or by electrophoretic separation. Alternatively, methods that amplify the hybridization probe to detectable levels can be used, such as Q β -replicase amplification. See, for example, Kramer, *et al.* *Nature*, 339: 401 (1989), Lizardi, *et al.* *Bio/Technology*, 6: 1197 (1988), and Lomell, *et al.*, *Clin. Chem.* 35: 1826 (1989) which are incorporated herein by reference.

In a preferred embodiment, amplification is by polymerase chain reaction using a pair of primers that flank and thereby amplify a selected deletion junction subsequence. Selection of primers is readily apparent to one of skill in the art using the sequence listings of the present invention. For example, a pair of PCR primers
5 5'-TCGACGATTGGCACAT-3' ($T_m=55^\circ\text{C}$) and 5'-TCCCTCCCTGTATTTGTAT-3' ($T_m=56^\circ\text{C}$) will amplify a 469 base pair sequence including the BCG Δ 1a deletion junction, while 5'-CGTTCTTCGGAGGTTTC-3' ($T_m=56^\circ\text{C}$) and 5'-GGCGGCTGGGTGGA-3' ($T_m=60^\circ\text{C}$) will amplify a 471 base pair sequence including the BCG Δ 1b deletion junction.

F) Detection of Deletions through Detection of Expression Products of Deletion Sequences

In addition to the detection of deletions by the detection of either the deletion junction sequences or the deletion sequences, one may detect the absence of the
15 deletion by detecting the expression products of the deletion sequences. Thus, for example, where the deletion sequences express a protein, the presence of that protein indicates the absence of the deletion and thus is indicative of a virulent (non BCG-like) phenotype. Such proteins are referred to herein as "deletion polypeptides".

Means of determining proteins expressed by particular nucleic acid
20 sequences are well known to those of skill in the art. Typically this involves determining the longest open reading frame. This may be aided by the identification of initiation sites (e.g. ribosome binding sites). The protein encoded by the largest open reading frame is determined using codon preferences for the specific organism from which the nucleic acid is obtained. The polypeptide sequence listing may then be compared against a
25 sequence database, e.g. GenBank, to determine other sequences sharing substantial sequence identity with the calculated sequence. The expression of the protein may be verified by isolating and then sequencing proteins having the predicted length and charge characteristics.

Once deletion polypeptides are identified they may be detected by routine
30 methods well known to those of skill in the art. Typically this involves isolating and then detecting the polypeptide. The polypeptide may be isolated by a number of means well known to those of skill in the art. This includes typical methods of protein

purification such as high performance liquid chromatography (HPLC), electrophoresis, capillary electrophoresis, hyperdiffusion chromatography, thin layer chromatography, and the like. Methods of purifying and detecting proteins are well known to those of skill in the art (see, e.g., *Methods in Enzymology Vol. 182: Guide to Protein Purification*, M. Deutscher, ed. Vol. 182 (1990), which is incorporated herein by reference).

Alternatively, deletion polypeptides sequences may be detected using immunoassays utilizing antibodies specific for the deletion polypeptides. The production of such antibodies and their use in immunoassays is detailed below.

10 G) Antibodies to Deletion Polypeptides

Antibodies can be raised to the polypeptides encoded by the nucleic acids corresponding to the open reading frames present in the deletion regions of the present invention (deletion polypeptides). As used herein "antibodies" include immunoglobulin or a population of immunoglobins which specifically bind to an antigen. Thus an antibody may be monoclonal or polyclonal including individual, allelic, strain, or species variants, and fragments thereof, both in their naturally occurring (full-length) forms and in recombinant forms. Additionally, antibodies can be raised to these polypeptides in either their native configurations or in non-native configurations. Anti-idiotypic antibodies may also be used.

20

1) Antibody Production

A number of immunogens may be used to produce antibodies specifically reactive with deletion polypeptides. Recombinant polypeptides are the preferred immunogen for the production of monoclonal or polyclonal antibodies. Naturally occurring polypeptides may also be used either in pure or impure form. Synthetic peptides made using sequences described herein may also be used as immunogens for the production of antibodies.

Recombinant polypeptides are expressed in eukaryotic or prokaryotic cells and purified using standard techniques. The polypeptide is injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies may be generated for subsequent use in immunoassays to measure the presence and quantity of the polypeptide.

30

Methods of producing polyclonal antibodies are known to those of skill in the art. In brief, an immunogen, preferably a purified deletion polypeptide is mixed with an adjuvant and animals are immunized with the mixture. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to the polypeptide of interest. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the polypeptide is performed where desired. See, e.g., Coligan (1991) *Current Protocols in Immunology* Wiley/Greene, NY; and Harlow and Lane (1989) *Antibodies: A Laboratory Manual* Cold Spring Harbor Press, NY, which are incorporated herein by reference.

Monoclonal antibodies may be obtained by various techniques familiar to those skilled in the art. Description of techniques for preparing such monoclonal antibodies may be found in, e.g., Stites *et al.* (eds.) *Basic and Clinical Immunology* (4th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein; Harlow and Lane (1988) *Antibodies: A Laboratory Manual* CSH Press; Goding (1986) *Monoclonal Antibodies: Principles and Practice* (2d ed.) Academic Press, New York, NY; and particularly in Kohler and Milstein (1975) *Nature* 256: 495-497, which discusses one method of generating monoclonal antibodies.

Summarized briefly, this method involves injecting an animal with an immunogen. The animal is then sacrificed and cells taken from its spleen, which are then fused with myeloma cells (See, Kohler and Milstein (1976) *Eur. J. Immunol.* 6: 511-519, incorporated herein by reference). The result is a hybrid cell or "hybridoma" that is capable of reproducing *in vitro*.

Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells is enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences which encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse *et al.* (1989) *Science* 246: 1275-1281. In this manner, the individual antibody species obtained are the products of immortalized and cloned single B

cells from the immune animal generated in response to a specific site recognized on the immunogenic substance.

Other suitable techniques involve selection of libraries of antibodies in phage or similar vectors. See, Huse *et al. Science* 246: 1275-1281 (1989); and Ward, *et al. Nature* 341: 544-546 (1989). The polypeptides and antibodies of the present invention are used with or without modification, including chimeric antibodies. Frequently, the polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionucleotides, enzymes, substrates, cofactors, inhibitors, fluorescent moieties, chemiluminescent moieties, magnetic particles, and the like. Patents, teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241. Also, recombinant immunoglobulins may be produced. See, Cabilly, U.S. Patent No. 4,816,567; and Queen *et al. Proc. Nat'l Acad. Sci. USA* 86: 10029-10033 (1989).

Antibodies, including binding fragments and single chain versions, against predetermined fragments of deletion polypeptides can be raised by immunization of animals with conjugates of the fragments with carrier proteins as described above. Monoclonal antibodies are prepared from cells secreting the desired antibody. These antibodies can be screened for binding to normal or defective polypeptides, or screened for agonistic or antagonistic activity, *e.g.*, mediated through a receptor. These monoclonal antibodies will usually bind with at least a K_D of about 1 mM, more usually at least about 300 μ M, and most preferably at least about 0.1 μ M or better.

The antibodies of this invention can also be used for affinity chromatography in isolating deletion polypeptides. Columns can be prepared where the antibodies are linked to a solid support, *e.g.*, particles, such as agarose, Sephadex, or the like, where a bacterial lysate, or recombinant cell lysate is passed through the column, washed, and treated with increasing concentrations of a mild denaturant, whereby purified deletion polypeptides are released.

The antibodies can be used to screen expression libraries for particular expression products. Usually the antibodies in such a procedure will be labeled with a moiety allowing easy detection of presence of antigen by antibody binding.

In a preferred embodiment, antibodies to deletion polypeptides are used for the identification of cell populations expressing the polypeptides. By assaying the expression products of cells expressing the polypeptides it is possible to diagnose bacterial infections.

Antibodies raised against each polypeptide are useful to raise anti-idiotypic antibodies. These will be useful in detecting or diagnosing various immunological conditions related to the presence of the respective antigens.

2) Immunoassays

A particular deletion polypeptide can be measured by a variety of immunoassay methods. For a review of immunological and immunoassay procedures in general, see Stites and Terr (eds.) 1991 *Basic and Clinical Immunology* (7th ed.). Moreover, the immunoassays of the present invention can be performed in any of several configurations, e.g., those reviewed in Maggio (ed.) (1980) *Enzyme Immunoassay* CRC Press, Boca Raton, Florida; Tijan (1985) "Practice and Theory of Enzyme Immunoassays," *Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers B.V., Amsterdam; and Harlow and Lane *Antibodies, A Laboratory Manual*, *supra*, each of which is incorporated herein by reference. See also Chan (ed.) (1987) *Immunoassay: A Practical Guide* Academic Press, Orlando, FL; Price and Newman (eds.) (1991) *Principles and Practice of Immunoassays* Stockton Press, NY; and Ngo (ed.) (1988) *Non-isotopic Immunoassays* Plenum Press, NY.

Immunoassays for measurement of deletion polypeptides can be performed by a variety of methods known to those skilled in the art. In brief, immunoassays to measure the protein can be, e.g., competitive or noncompetitive binding assays. In competitive binding assays, the sample to be analyzed competes with a labeled analyte for specific binding sites on a capture agent bound to a solid surface. Preferably the capture agent is an antibody specifically reactive with a deletion polypeptide produced as described above. The concentration of labeled analyte bound to the capture agent is inversely proportional to the amount of free analyte present in the sample.

In a competitive binding immunoassay, the deletion polypeptide present in the sample competes with labelled protein for binding to a specific binding agent, for example, an antibody specifically reactive with a particular deletion polypeptide. The binding agent is, *e.g.*, bound to a solid surface to produce separation of bound labelled polypeptide from the unbound labelled polypeptide. Alternately, the competitive binding assay may be conducted in liquid phase and any of a variety of techniques known in the art may be used to separate the bound labelled protein from the unbound labelled protein. Following separation, the amount of bound labeled protein is determined. The amount of polypeptide present in the sample is inversely proportional to the amount of labelled polypeptide binding.

Alternatively, a homogenous immunoassay may be performed in which a separation step is not needed. In these immunoassays, the label on the protein is altered by the binding of the protein to its specific binding agent. This alteration in the labelled protein results in a decrease or increase in the signal emitted by label, so that measurement of the label at the end of the immunoassay allows for detection or quantitation of the polypeptide.

Deletion polypeptides may also be detected by a variety of noncompetitive immunoassay methods. For example, a two-site, solid phase sandwich immunoassay may be used. In this type of assay, a binding agent for the protein, for example an antibody, is attached to a solid support. A second protein binding agent, which is also an antibody, and which binds the protein at a different site, is labelled. After binding at both sites on the protein, the unbound labelled binding agent is removed and the labelled binding agent bound to the solid phase is measured. The amount of labelled binding agent bound is directly proportional to the amount of polypeptide in the sample.

Western blot analysis can be used to determine the presence of a deletion polypeptide in a sample. Electrophoresis is carried out, for example, on a bacterial sample suspected of containing the deletion polypeptide. Following electrophoresis to separate the proteins, and transfer of the proteins to a suitable solid support such as a nitrocellulose filter, the solid support is incubated with an antibody reactive with the protein. This antibody is labelled, or alternatively may be it is detected by subsequent incubation with a second labelled antibody that binds the primary antibody.

The immunoassay formats described above employ labelled assay components. The label can be in a variety of forms as described above. The choice of label depends on sensitivity required, ease of conjugation with the compound, stability requirements, and available instrumentation. For a review of various labelling or signal producing systems which may be used, see U.S. Patent No. 4,391,904, which is incorporated herein by reference.

Antibodies reactive with a particular protein can also be measured by a variety of immunoassay methods. For a review of immunological and immunoassay procedures applicable to the measurement of antibodies by immunoassay techniques, see Stites and Terr (eds.) *Basic and Clinical Immunology* (7th ed.) *supra*; Maggio (ed.) *Enzyme Immunoassay, supra*; and Harlow and Lane *Antibodies, A Laboratory Manual, supra*.

In brief, immunoassays to measure antisera reactive with polypeptides include competitive and noncompetitive binding assays. In competitive binding assays, the sample analyte competes with a labeled analyte for specific binding sites on a capture agent bound to a solid surface. Preferably the capture agent is a purified recombinant deletion polypeptide as described above. Other sources of polypeptides, including isolated or partially purified naturally occurring protein, can also be used. Noncompetitive assays are typically sandwich assays, in which the sample analyte is bound between two analyte-specific binding reagents. One of the binding agents is used as a capture agent and is bound to a solid surface. The second binding agent is labelled and is used to measure or detect the resultant complex by visual or instrument means. A number of combinations of capture agent and labelled binding agent can be used. A variety of different immunoassay formats, separation techniques and labels can be also be used similar to those described above for the measurement of deletion polypeptides.

II. Preparation of Deletion-Containing Mycobacteria

Mycobacteria containing specific deletions may be prepared by using methods of homologous recombination well known to those of skill in the art. In brief, homologous recombination is a natural cellular process which results in the scission of two nucleic acid molecules having identical or substantially similar (*i.e.* "homologous") sequences, and the ligation of the two molecules such that one region of each initially

present molecule is now ligated to a region of the other initially present molecule (Sedivy, *Bio/Technol.*, 6: 1192-1196 (1988)).

Homologous recombination is exploited by a number of various methods of "gene targeting" well known to those of skill in the art. (see, for example, Mansour *et al. Nature*, 336: 348-352 (1988); Capecchi *Trends Genet.* 5: 70-76 (1989); Capecchi *Science* 244: 1288-1292 (1989); Capecchi *et al.* pages 45-52 In: *Current Communications in Molecular Biology*, Capecchi, M.R. (ed.), Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1989); Frohman *et al. Cell* 56: 145-147 (1989)). Some approaches focus on increasing the frequency of recombination between two DNA molecules by treating the introduced DNA with agents which stimulate recombination (e.g. trimethylpsoralen, UV light, etc.), however, most approaches utilize various combinations of selectable markers to facilitate isolation of the transformed cells.

One such selection method is termed positive/negative selection (PNS) (Thomas and Capecchi *Cell* 51: 503-512 (1987)). This method involves the use of two selectable markers: one a positive selection marker such as the bacterial gene for neomycin resistance (*neo^r*); the other a negative selection marker such as the herpes virus thymidine kinase (*tk*) gene. *Neo^r* confers resistance to the drug G-418, while herpes *tk* renders cells sensitive to the nucleoside analog gancyclovir (GANC) or 1-(2-deoxy-2-fluoro-b-d-arabinofuranosyl)-5-iodouracil (FIAU). The DNA encoding the positive selection marker in the transgene (e.g. *neo^r*) is generally linked to an expression regulation sequence that allows for its independent transcription in mycobacteria. It is flanked by first and second sequence portions of at least a part of the deletion or deletion flanking sequences.

These first and second sequence portions target the transgene to a specific nucleotide sequence. A second independent expression unit capable of producing the expression product for a negative selection marker, e.g. for herpes virus *tk* is positioned adjacent to or in close proximity to the distal end of the first or second portions of the first DNA sequence. Upon transfection, some of the mycobacteria incorporate the transgene by random integration, others by homologous recombination between the endogenous allele and sequences in the transgene. As a result, one copy of the targeted nucleic acid is disrupted by homologous recombination with the-transgene with simultaneous loss of the sequence encoding herpes *tk* gene. Random integrants, which

occur via the ends of the transgene, contain herpes tk and remain sensitive to GANC or FIAU. Therefore, selection, either sequentially or simultaneously with G418 and GANC enriches for transfected mycobacteria containing the transgene integrated into the genome by homologous recombination.

5 Methods of homologous recombination in mycobacteria are described in greater detail by Ganjam *et al.* *Proc. Natl. Acad. Sci. USA*, 88: 5433-5437 (1991) and Aldovini *et al.*, *J. Bacteriol.*, 175: 7282-7289 (1993) which are incorporated herein by reference.

10 **III. Screening for Drug Susceptibility/Therapeutics**

 The expression products of the open reading frames in the BCG Δ 1, BCG Δ 2, and BCG Δ 3 deletions of the present invention are targets for anti-mycobacterial drugs. To determine particularly suitable drug targets, open reading frames and surrounding expression control sequences are introduced into avirulent strains of mycobacteria, alone or in combination with other open reading frame regions to
15 determine which regions are critical for virulence. Once particular genes are identified as critical for virulence, anti-mycobacterial agents are designed to inhibit expression of the critical genes, or to attack the critical gene products. For instance, antibodies are generated against the critical gene products and used as prophylactic or therapeutic
20 agents. Alternatively, small molecules can be screened for the ability to selectively inhibit expression of the critical gene products, *e.g.*, using recombinant expression systems which include the gene's endogenous promoter. These small molecules are then used as therapeutics, or prophylactic agents to inhibit mycobacterial virulence.

 In another embodiment, anti-mycobacterial agents which render a virulent
25 mycobacterium avirulent can be operably linked to expression control sequences and used to transform a virulent mycobacterium. Such anti-mycobacterial agents inhibit the replication of a specified mycobacterium upon transcription or translation of the agent in the mycobacterium.

 Such transformed mycobacteria are useful as vaccine components, and as
30 components of immunological infectivity assays. For instance, an animal's blood can be monitored for the presence of anti-mycobacterial antibodies using the procedures described herein, using transformed avirulent mycobacterial components in various

immunological assays. Anti-mycobacterial agents useful in this invention include, without limitation, antisense genes, ribozymes, decoy genes, transdominant proteins and suicide genes.

5 An antisense nucleic acid is a nucleic acid that, upon expression, hybridizes to a particular mRNA molecule, to a transcriptional promoter or to the sense strand of a gene. By hybridizing, the antisense nucleic acid interferes with the transcription of a complementary DNA, the translation of an mRNA, or the function of a catalytic RNA. Antisense molecules useful in this invention include those that hybridize to gene transcripts in the region of the deletions of the invention, particularly deletion
10 region 1.

A ribozyme is a catalytic RNA molecule that cleaves other RNA molecules having particular nucleic acid sequences. Ribozymes useful in this invention are those that cleave deletion gene transcripts. Examples include hairpin and hammerhead ribozymes.

15 A decoy nucleic acid is a nucleic acid having a sequence recognized by a regulatory DNA binding protein (*i.e.*, a transcription factor). Upon expression, the transcription factor binds to the decoy nucleic acid, rather than to its natural target in the genome. Useful decoy nucleic acid sequences include any sequence to which a transcription factor binds in the deletion regions of the present invention.

20 A transdominant protein is a protein whose phenotype, when supplied by transcomplementation, will overcome the effect of the native form of the protein. For instance, an avirulent mycobacterium can be rendered virulent by introducing transdominant proteins from deletion region 1.

25 A suicide gene produces a product which is cytotoxic. In the vectors of the present invention, a suicide gene is operably linked to an inducible expression control sequences which is stimulated upon infection of a cell by a mycobacterium.

IV. Use of Expressed "Deletion Proteins" in a Vaccine

30 The deletion polypeptides encoded by the open reading frames in BCG Δ 1, BCG Δ 2, and BCG Δ 3 may be recombinantly expressed and used as components of immunological assays as described above or in vaccines. Expression of polypeptides

encoded by the open reading frames of the BCG Δ 1, BCG Δ 2, or BCG Δ 3 deletions may be accomplished by means well known to those of skill in the art.

In brief, the expression of natural or synthetic nucleic acids encoding deletion polypeptides will typically be achieved by operably linking the DNA or cDNA to a promoter (which is either constitutive or inducible), followed by incorporation into an expression vector. The vectors can be suitable for replication and integration in either prokaryotes or eukaryotes. Typical expression vectors contain transcription and translation terminators, initiation sequences, and promoters useful for regulation of the expression of polynucleotide sequence encoding deletion polypeptides.

To obtain high level expression of a cloned gene, such as those polynucleotide sequences encoding deletion polypeptides, it is desirable to construct expression plasmids which contain, at the minimum, a promoter to direct transcription, a ribosome binding site for translational initiation, and a transcription/translation terminator. The expression vectors may also comprise generic expression cassettes containing at least one independent terminator sequence, sequences permitting replication of the plasmid in both eukaryotes and prokaryotes, *i.e.*, shuttle vectors, and selection markers for both prokaryotic and eukaryotic systems. For detailed techniques employed in the recombinant expression of deletion proteins *see*, for example, Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual* (2nd Ed., Vols. 1-3, Cold Spring Harbor Laboratory (1989)), *Methods in Enzymology*, Vol. 152: *Guide to Molecular Cloning Techniques* (Berger and Kimmel (eds.), San Diego: Academic Press, Inc. (1987)), or *Current Protocols in Molecular Biology*, (Ausubel, *et al.* (eds.), Greene Publishing and Wiley-Interscience, New York (1987), all of which are incorporated herein by reference.

The expressed deletion polypeptides may be used in a variety of assays. For example, the deletion polypeptides can be used as reagents in immunoblot assays to test whether a patient was previously exposed to virulent mycobacteria (*i.e.*, to test whether the patient has antibodies to the deletion polypeptide). These assays have the advantage of discriminating between previous exposure to an avirulent mycobacterium (*e.g.*, one used in a vaccine) and exposure to a virulent mycobacterium. Thus, vaccinated individuals can be tested for antibodies to the virulent mycobacterium without regard to whether the patient has been vaccinated with an avirulent mycobacterium.

The deletion polypeptides can also be used as antigenic vaccine components to direct antibodies to elements which are critical for virulence. These polypeptides can be added to existing vaccines (*e.g.*, those based upon avirulent mycobacteria and which lack the deletion polypeptide) to supplement the range of antigenicity conferred by the vaccine, or they may be used apart from other mycobacterial antigens. The vaccines of the invention contain as an active ingredient an immunogenically effective amount of a deletion polypeptide or of a recombinant vector which includes the deletion polypeptide. The immune response can include the generation of antibodies; activation of cytotoxic T lymphocytes (CTL) against cells presenting peptides derived from the polypeptides or other mechanisms well known in the art. See *e.g.* Paul *Fundamental Immunology Third Edition* published by Raven press New York (incorporated herein by reference) for a description of immune response. Useful carriers are well known in the art, and include, for example, thyroglobulin, albumins such as human serum albumin, tetanus toxoid, and polyamino acids such as poly(D-lysine:D-glutamic acid). The vaccines can also contain a physiologically tolerable (acceptable) diluent such as water, phosphate buffered saline, and further typically include an adjuvant. Adjuvants such as incomplete Freund's adjuvant, aluminum phosphate, aluminum hydroxide, or alum are materials well known in the art.

The compositions are suitable for single administrations or a series of administrations. When given as a series, inoculations subsequent to the initial administration are given to boost the immune response and are typically referred to as booster inoculations.

The vaccine compositions of the invention are intended for parenteral, topical, oral or local administration. Preferably, the pharmaceutical compositions are administered parenterally, *e.g.*, intravenously, subcutaneously, intradermally, or intramuscularly. Thus, the invention provides compositions for parenteral administration that comprise a solution of the agents described above dissolved or suspended in an acceptable carrier, preferably an aqueous carrier. A variety of aqueous carriers may be used, *e.g.*, water, buffered water, 0.4% saline, 0.3% glycine, hyaluronic acid and the like. These compositions may be sterilized by conventional, well known sterilization techniques, or may be sterile filtered. The resulting aqueous solutions may be packaged for use as is, or lyophilized, the lyophilized preparation being combined with a sterile

solution prior to administration. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions, such as pH adjusting and buffering agents, tonicity adjusting agents, wetting agents and the like, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, etc.

For solid compositions, conventional nontoxic solid carriers may be used which include, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharin, talcum, cellulose, glucose, sucrose, magnesium carbonate, and the like. For oral administration, a pharmaceutically acceptable nontoxic composition is formed by incorporating any of the normally employed excipients, such as those carriers previously listed, and generally 10-95% of active ingredient and more preferably at a concentration of 25%-75%.

For aerosol administration, the polypeptides are preferably supplied in finely divided form along with a surfactant and propellant. The surfactant should be nontoxic, and preferably soluble in the propellant. Representative of such agents are the esters or partial esters of fatty acids containing from 6 to 22 carbon atoms, such as caproic, octanoic, lauric, palmitic, stearic, linoleic, linolenic, olesteric and oleic acids with an aliphatic polyhydric alcohol or its cyclic anhydride. Mixed esters, such as mixed or natural glycerides may be employed. A carrier can also be included, as desired, as with, *e.g.*, lecithin for intranasal delivery.

The amount of vaccine administered to the patient will vary depending upon the composition being administered, the physiological state of the patient and the manner of administration.

Live attenuated recombinant viruses which include the deletion polypeptide, such as recombinant vaccinia or adenovirus vectors, are convenient alternatives as vaccines because they are inexpensive to produce and are easily transported and administered. Vaccinia vectors and methods useful in immunization protocols are described, for example, in U.S. Patent No. 4,722,848, incorporated herein by reference.

Deletion sequences and subsequences of this invention may also be used in methods of genetic immunization. Briefly, genetic immunization involves transfecting

cells *in vivo* with nucleic acids encoding pathogen specific antigens. The transformed host cells then express the antigen thereby stimulating the host immune system.

In the present invention, antigen-encoding deletion region sequences are used to transform mammalian host cells thereby resulting in the expression of the antigen by the host. This provokes an immune response by the host against the expressed antigen thereby conferring immunity on the host. Methods of genetic immunization are well known to those of skill in the art (see, *e.g.*, Wang *et al. Proc. Natl. Acad. Sci. USA*, 90: 4156-4160 (1993); Ulmer *et al., Science*, 259: 1745-1749 (1993); Fynan *et al. DNA Cell Biol.*, 12: 785-789 (1993); Fynan *et al. Proc. Natl. Acad. Sci. USA*, 90: 11478-11482 (1993); Robinson *et al. Vaccine*, 11: 957-960 (1993); and Martinon *et al. Eur. J. Immunol.*, 23: 1719-1722 (1993), which are incorporated herein by reference.

VI. Use of Promoters within Deletion Sequences for Expression of Recombinant Proteins

Bacille Calmette-Guérin (BCG) contains all three deletions (BCG Δ 1, BCG Δ 2, and BCG Δ 3) and yet is able to grow and reproduce indicating that the sequences contained within the deletion are not essential for bacterial viability. These deletion regions therefore make good target sites for the insertion of heterologous DNA as mycobacteria are tolerant of disruption of the native genome in these regions. The BCG Δ 1, BCG Δ 2, and BCG Δ 3 deletion regions therefore provide suitable target sites for the incorporation of expression cassettes and the subsequent expression of exogenous gene products. The expression cassettes typically comprise a nucleic acid sequence under the control of a promoter. The promoter may be either constitutive or inducible. The cassette may additionally comprise a selectable marker such as an antibiotic resistance gene, a gene encoding a fluorescent marker (*e.g.* green fluorescent protein), or a gene encoding an enzymatic marker (*e.g.* β -galactosidase).

Alternatively, genes under the control of endogenous promoters may be used as well. In one embodiment, reporter genes under the control of endogenous promoters found within the deletion sequences may be inserted at the deletion sites. These reporter genes may be utilized as an assay for antimycobacterial compounds that act by inhibiting transcription or translation of deletion sequences. Assaying for the

reporter gene product in the presence of an antimycobacterial compound provides a measure of efficacy of that compound in upregulating or downregulating deletion sequence genes. Methods of use of mycobacterial reporter gene assays to screen for drug activity are described by Cooksey *et al.*, *Antimicrob. Agents Chemother.*, 37: 1348-1352 (1993), and Jacobs *et al.*, *Science*, 260: 819-822 (1993) which are incorporated herein by reference.

EXAMPLES

The following examples are offered by way of illustration, not by way of limitation.

Example 1

Identification of Virulence-Attenuating Deletions

Bacterial Culture

All strains of Mycobacteria used in this study were maintained in 7H9 (Difco, Detroit Michigan, USA) media supplemented with OADC (BBL) or were grown on 7H11 agar supplemented with oleic acid albumin dextrose complex (OADC). *Escherichia coli* (strain DH5 α or NM554) was used as a host for all recombinant plasmids and cosmids. *E. coli* was maintained in LB medium with or without agar. Carbenicillin (100 μ g/ml) was used in place of ampicillin for the selection of all *E. coli* plasmids.

Extraction of High Molecular Weight DNA

High molecular weight chromosomal DNA was prepared by diluting a late log phase culture of the respective mycobacterium 1:10 into a liter of 7H9 medium containing 1.5% glycine and continuing growth for 4 to 5 days. The cells were then harvested by centrifugation, washed once in TE (pH 8.0) and resuspended in 4 ml of 25% sucrose in 10X TE. 100 μ g of lysozyme was added and the preparation was incubated at 37°C for 2 hr followed by the addition of 100 μ g of proteinase K and sarkosyl to a concentration of 1% weight/volume. Following overnight incubation at 65°C the mixture was extracted 4 times with chloroform isoamyl alcohol 24:1, once with phenol/chloroform (1:1), and twice again with chloroform isoamyl alcohol. The resulting high molecular weight DNA was then run on a CsCl gradient as described by

Hull *et al. Infect. Immun.*, 33: 933-938 (1981), which is incorporated herein by reference, and subsequently dialyzed against 4 changes of TE. BCG DNA was physically sheared by passage through a 22 gauge needle until an average size of 3-10 kb was obtained (20-25 passages). This DNA was then biotinylated using photobiotin (Clonetech, Palo Alto, California, USA) according to the method of Straus and Ausubel, *Proc. Natl. Acad. Sci. USA*, 87: 1889-1893 (1990), which is incorporated herein by reference.

DNA Subtraction

DNA subtraction was carried out between virulent *M. tuberculosis* H37Rv and avirulent BCG. H37Rv chromosomal DNA was selected because it was the most readily available chromosomal DNA from a virulent strain. In addition, *M. bovis* and *M. tuberculosis* H37Rv are highly homologous.

M. bovis/*M. tuberculosis* specific probes were generated by the method of Straus and Ausubel, *supra*. with the following modifications. Sheared and biotinylated BCG DNA was used in a 10:1 excess for each round of subtraction. Wild type *M. tuberculosis* H37Rv DNA was digested with Sau3A to an average size of 1 kb. Hybridization conditions were 1M NaCl and 65 °C for 18 hours. Following five cycles (successive denaturation and reassociations) of subtraction, Sau3A1 adaptors (GAACTCTCGAGACATCACCGTCC and GATCGGACGGTGATGTCTCGAGAGTG) were ligated to the subtraction product and amplified in a PCR reaction for 35 cycles (30 sec at 95°C, 30 sec at 55°C, and 3 min at 72°C). The *M. tuberculosis*/*M. bovis* specific probes were radiolabeled by using one strand of the adaptor (GAACTCTCGAGACATCACCGTCC) as a primer and labeling with ³²P dCTP using the Klenow fragment of DNA polymerase.

An *M. bovis* cosmid library was constructed in the BamHI site of sCOS (Stratagene, La Jolla California, USA) with subsequent *in vitro* packaging and infection of *E. coli* strain NM554 (Stratagene). 600 colonies were picked to Nytran circular membranes and the membranes prepared according to the method of Grunstein and Hogness, *Proc. Natl. Acad. Sci. USA*, 72: 3961 (1975), which is incorporated herein by reference. These filters were then probed using the BCG subtracted probe and positive clones selected for further analysis. Cosmid DNA was prepared from selected clones by the method of Birnboim and Doly, *Nucleic Acids. Res.*, 7: 1513 (1973) which is

incorporated herein by reference. Restriction fragments that hybridize with the MTB/MBV specific probe were further subcloned into pGEM7z or pGEM5z (Promega, Madison, Wisconsin, USA) for deletion analysis.

Plasmid DNA for DNA sequencing was prepared using Qiagen minicolumns (Qiagen Inc. Chatsworth California, USA) and sequenced by the method of Henikoff, *Gene*, 28: 351-359 (1984), which is incorporated herein by reference, using the Erase A Base System (Promega). DNA sequencing reactions were run using a Perkin Elmer 9600 thermocycler and analyzed on an automated ABI sequencer. Analysis and assembly of contiguous DNA sequence was done using the ABI analysis software and SeQuencher sequence analysis software by Gene Clones Corp (Ann Arbor, Michigan, USA).

Deletion Region 1 (BCGA1)

Sequence analysis of over 16 kb of MBV region 1 and homologous regions in BCG revealed the precise junctions for the deletion in BCG. Eight open reading frames were identified with codon usage biases matching that of known MTB and MBV genes (see map Figure 4). The potential start and stop codons and predicted maximum protein coding capacity are listed in Figure 4. Consensus ribosomal binding site sequences were found near potential start codons for seven of eight open reading frames. TBLASTN and FASTA sequence homology analysis with each potential ORF-encoded protein revealed significant homologies for 3 of 8 open reading frames in region 1.

Most notable is the ORF1C homology to an unpublished and uncharacterized sequence listed in Genbank as *M. tuberculosis* antigen esat6. A 65 base pair repeated overlapping (repeated ~2 1/2 times) sequence was also recognized within the ORF1C (esat6) open reading frame. Also noteworthy are the significant homologies identified between ORF1H and bacterial serine proteases including *B. subtilis* subtilisin. Of the eight recognized open reading frames, four (ORFs 1B, 1C, 1D, and 1E) are located entirely within the 9 kb region deleted in BCG. One ORF traverses the BCG deletion junction in virulent *M. bovis*.

DNA probes from the 9 kb deletion in region 1 demonstrated that this region is absent in all BCG substrains and present in all virulent MBV and MTB strains tested. Furthermore, restriction fragment patterns observed in Southern blot analysis

with region 1 probes are non-polymorphic and identical in virulent MBV and MTB. This region has far fewer direct and indirect repeats than the regions 2 (BCG Δ 2) and 3 (BCG Δ 3) characterized below.

5 The sequence of a small region, estimated to be less than 20 bp between basepair coordinates 10654 and 10664 in region 1 has been recalcitrant to automated sequencing. Therefore, pending sequence confirmation, the base pair coordinates given in the region 1 map (Figure 4) are approximations. The precise sequence determination is likely to effect the Orf1E open reading frame.

10 Deletion Region 2 (BCG Δ 2)

Sequence analysis of over 15 kb of MBV region 2 and homologous regions in BCG revealed the precise junctions for an 11 kb deletion in BCG. Thirteen open reading frames were identified with codon usage biases matching that of known MTB and MBV genes (see map Figure 5). The potential start and stop codons and predicted maximum protein coding capacity are also shown in Figure 5. Candidate consensus sequences resembling ribosomal binding sites were found near potential start codons for eight open reading frames. Of the thirteen open reading frames recognized in BCG Δ 2, nine are located entirely within the 11 kb region deleted in most BCG strains while ORF2B2 and ORF2I traverse the deletion junctions.

20 TBLASTN and FASTA sequence homology analysis with each potential ORF-encoded protein revealed significant homologies for five open reading frames in BCG Δ 2. A protein encoded by ORF2C exhibits striking similarity to the *E. coli* *iciA* protein which is thought to play a role in inhibiting and regulating the initiation of chromosomal replication. The *iciA* protein product is a member of the large LysR family of transcriptional regulatory proteins. Orf2F is highly homologous to an *S. typhimurium* ribonucleotide diphosphate reductase and a region of the *E. coli* and *S. typhimurium* proUVWX operon. Orf2H was found to have significant homology to *E. coli* and *S. typhimurium* permeases involved in aromatic amino acid transport and a eukaryotic cell retroviral receptor.

30 The Orf2G encoded protein was identical to the MTB *mpt64* gene previously thought to encode a secreted antigen which is specifically expressed by MTB

and not BCG strains. Recent analysis of mpt64 expression revealed that three BCG substrains do express mpt64 (Moreau, Tokyo, Russian). Probes specific for mpt64 or other non-repetitive parts of region 2 hybridized to all MTB strains tested and the same three BCG substrains shown to express mpt64. Of interest is the finding that these three BCG substrains are derived from the original Pasteur strain prior to 1925. The current Pasteur strain and all strains derived from the original Pasteur strain after 1925, including the Connaught strain used in the subtractive analysis in this study, are deleted in the 11 kb DNA segment contained within BCG Δ 2. These data indicate that an additional mutational event deleting the 11 kb segment of region 2, occurred in the BCG Pasteur strain sometime after 1925.

Southern blot analysis with probes from different segments of region 2 revealed a repetitive element located within a 2 kb segment (8-10 kb) of region 2. This repetitive element is ubiquitous in all tubercle bacilli tested. This element provides a marker suitable for RFLP analysis of mycobacterial strains.

Deletion Region 3 (BCG Δ 3)

Sequence analysis of the almost 11 kb region 3 sequence and comparison to a homologous region in BCG precisely identified the deletion junctions for BCG. Twelve potential open reading frames were recognized in the region 3 sequence, seven of which are entirely located within the 9 kb region deleted in BCG. At least 9 ORFs in BCG Δ 3 exhibit codon usage preferences comparable to that of the tubercle bacilli. Sequence homology analysis of presumptive protein sequences encoded by six open reading frames in region 3 revealed highly significant homology to listed sequences. Orfs3B, 3D, and 3E exhibit homology to phage sequences, suggesting a phage derivation for 4 or more kb of DNA in region 3. Homology to putative open reading frames in two *M. leprae* cosmids was also observed including homology to a putative *bid* gene encoding a protein involved in biotin synthesis. Also of interest was homology between ORF3A and an MTB sequence (*mce*) associated with cell invasion and intracellular survival.

Southern blot analysis with segments of region 3 deleted in BCG revealed that prototype lab strains of virulent MBV and MTB all carry deletion region 3 DNA. However, clinical isolates from PHRI are highly polymorphic or deleted in region 3.

This region contains many large direct and indirect repeats and, as mentioned above, at least 2 ORFs are homologous to phage sequences including homology to DNA invertases or recombinases. The repetitive nature of this region and the possible presence of a DNA recombinase could explain the polymorphisms observed in this region.

5 The sequence of a small region, estimated to be much less than 200 bp and located close to 9400 bp in Figure 3, was recalcitrant to automated sequencing and remains to be determined. Therefore, the base pair coordinates given in the region 3 map (Figure 6) 3' to the 9kb marker are approximations. The precise sequence determination of region is likely to effect the length of open reading frames 3H and 3L.

10 The foregoing subtractive analysis identified 3 regions in virulent *M. bovis* and *M. tuberculosis* prototype strains which are deleted in the avirulent BCG strain. The deletion located in region 2 may not have arisen in the original BCG Pasteur strain as this region is only deleted in strains derived from the original Pasteur strain after 1925. Region 3 is present in virulent MTB and MBV lab prototype strains (H37Rv, Erdman)
15 and is highly polymorphic and at least partially deleted in the majority of MTB clinical isolates tested. Region 1 is apparently conserved and intact in all virulent MBV and MTB strains tested to date while all avirulent BCG strains tested to date are missing approximately 9kb from region 1.

Example 2

20 Screening and Identification of an Avirulent Mycobacterium

 The ³² P labeled subtraction probe obtained in Example 1, was used to probe EcoRI and BamHI restricted chromosomal DNAs from BCG Connaught, *Mycobacterium bovis*, and various strains of *Mycobacterium tuberculosis* in a Southern blot. The hybridization was performed at 70°C in 6X SSC overnight.

25 The resulting Southern blot is illustrated in Figure 8. The probe showed no labeling of BCG reflecting the presence of all three deletions, while the other strains were labeled.

 The above examples are provided to illustrate the invention but not to limit
30 its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference.

WHAT IS CLAIMED IS:

1 1. A marker for an avirulent mycobacterium, said marker comprising
2 a first nucleic acid that specifically hybridizes under stringent conditions with a second
3 nucleic acid or a complement of said second nucleic acid where said second nucleic acid
4 or complement of said second nucleic acid is selected from the group consisting of
5 BCGΔ1a, BCGΔ1b, BCGΔ2a, BCGΔ2b, BCGΔ3a, BCGΔ3b, BCGΔ1ab, BCGΔ2ab,
6 BCGΔ3ab, BCGΔ1, BCGΔ2, and BCGΔ3.

7 2. The marker of claim 1, wherein said marker specifically hybridizes
8 under stringent conditions to a nucleic acid from BCG, but not to a nucleic acid from
9 *Mycobacterium tuberculosis* or *Mycobacterium bovis*, or where said marker specifically
10 hybridizes under stringent conditions to a nucleic acid from *Mycobacterium tuberculosis*
11 or *Mycobacterium bovis*, but not to a nucleic acid from BCG.

1 2 3. The marker of claim 2, wherein said marker comprises a
3 subsequence of a nucleic acid where said nucleic acid is selected from the group
4 consisting of BCGΔ1a, BCGΔ1b, BCGΔ2a, BCGΔ2b, BCGΔ3a, BCGΔ3b, BCGΔ1ab,
5 BCGΔ2ab, BCGΔ3ab, BCGΔ1, BCGΔ2, and BCGΔ3.

1 4. The marker of claim 2, wherein said marker is selected from the
2 group consisting of BCGΔ1a, BCGΔ1b, BCGΔ2a, BCGΔ2b, BCGΔ3a, BCGΔ3b,
3 BCGΔ1ab, BCGΔ2ab, BCGΔ3ab, BCGΔ1, BCGΔ2, and BCGΔ3.

1 5. The marker of claim 2, wherein said marker comprises a nucleic
2 acid having at least 90 percent sequence identity with a sequence selected from the group
3 consisting of BCGΔ1a, BCGΔ1b, BCGΔ2a, BCGΔ2b, BCGΔ3a, BCGΔ3b, BCGΔ1ab,
4 BCGΔ2ab, BCGΔ3ab, BCGΔ1, BCGΔ2, and BCGΔ3.

1 6. The marker of claim 2, wherein said marker comprises a
2 radioactive nucleotide probe.

1 7. The marker of claim 2, wherein said subsequence is a sequence
2 selected from an open reading frame of a deletion, said deletion being selected from the
3 group consisting of BCG Δ 1, BCG Δ 2, BCG Δ 3.

1 8. A polypeptide encoded by a subsequence of a deletion sequence
2 selected from the group consisting of BCG Δ 1, BCG Δ 2, and BCG Δ 3.

1 9. The polypeptide of claim 8, wherein the subsequence is selected
2 from an open reading frame (ORF) of a deletion, said deletion being selected from the
3 group consisting of BCG Δ 1, BCG Δ 2, BCG Δ 3.

1 10. An antibody that binds specifically to the polypeptide of claim 8.

1 11. A recombinant cell comprising a first nucleic acid that hybridizes
2 under stringent conditions with a second nucleic acid or a complement of said second
3 nucleic acid where said second nucleic acid or complement of said second nucleic acid is
4 selected from the group consisting of BCG Δ 1a, BCG Δ 1b, BCG Δ 2a, BCG Δ 2b, BCG Δ 3a,
5 BCG Δ 3b, BCG Δ 1ab, BCG Δ 2ab, BCG Δ 3ab, BCG Δ 1, BCG Δ 2, and BCG Δ 3.

1 12. The recombinant cell of claim 11, wherein the cell is a
2 Mycobacterium.

1 13. The cell of claim 11, wherein the cell expresses a polypeptide
2 encoded by an intact open reading frame from BCG Δ 1, BCG Δ 2, and BCG Δ 3.

1 14. The cell of claim 11, wherein said cell is a mycobacterium having
2 one or more deletions in the genomic regions selected from the group consisting of
3 BCG Δ 1, BCG Δ 2, and BCG Δ 3, wherein said deletions result in the attenuation of an
4 otherwise virulent strain of mycobacterium and wherein said deletions are present in up
5 to two of said regions.

1 15. The mycobacterium of claim 14, wherein said deletions comprise a
2 deletion selected from the group consisting of BCG Δ 1, BCG Δ 2, and BCG Δ 3.

1 16. A method of distinguishing between an attenuated and a virulent
2 mycobacterium, said method comprising detecting the presence or absence of a first
3 nucleic acid that hybridizes under stringent conditions with a second nucleic acid or a
4 complement of said second nucleic acid where said second nucleic acid or complement of
5 said second nucleic acid is selected from the group consisting of BCG Δ 1a, BCG Δ 1b,
6 BCG Δ 2a, BCG Δ 2b, BCG Δ 3a, BCG Δ 3b, BCG Δ 1ab, BCG Δ 2ab, BCG Δ 3ab, BCG Δ 1,
7 BCG Δ 2, and BCG Δ 3.

1 17. The method of claim 16, wherein said first nucleic acid specifically
2 hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic
3 acid from *Mycobacterium tuberculosis* or *Mycobacterium bovis*, or where said first
4 nucleic acid specifically hybridizes under stringent conditions to a nucleic acid from
5 *Mycobacterium tuberculosis* or *Mycobacterium bovis*, but not to a nucleic acid from
6 BCG.

1 18. The method of claim 17, wherein said first sequence is amplified
2 prior to detection.

1 19. The method of claim 17, wherein said first sequence is amplified
2 by the polymerase chain reaction.

1 20. A method of claim 17, wherein said detecting comprises a Southern
2 blot.

1 21. A method of claim 17, wherein said detecting comprises detecting a
2 polypeptide encoded by said first nucleic acid.

2 22. The method of claim 21, wherein the polypeptide is encoded by an
3 intact open reading frame of a nucleotide sequence selected from the group consisting of
4 BCG Δ 1, BCG Δ 2, and BCG Δ 3.

1 23. The method of claim 21, wherein the polypeptide is visualized by
2 antibody hybridization.

1 24. A method for determining whether an attenuated or a virulent
2 Mycobacterium is present in a sample comprising:
3 providing a first nucleic acid that hybridizes under stringent conditions
4 with a second nucleic acid or a complement of said second nucleic acid where said
5 second nucleic acid or complement of said second nucleic acid is selected from the group
6 consisting of BCG Δ 1a, BCG Δ 1b, BCG Δ 2a, BCG Δ 2b, BCG Δ 3a, BCG Δ 3b, BCG Δ 1ab,
7 BCG Δ 2ab, BCG Δ 3ab, BCG Δ 1, BCG Δ 2, and BCG Δ 3; and
8 hybridizing said first nucleic acid to the biological sample.

1 25. The method of claim 24, wherein said first nucleic acid specifically
2 hybridizes under stringent conditions to a nucleic acid from BCG, but not to a nucleic
3 acid from *Mycobacterium tuberculosis* or *Mycobacterium bovis*, or where said first
4 nucleic acid specifically hybridizes under stringent conditions to a nucleic acid from
5 *Mycobacterium tuberculosis* or *Mycobacterium bovis*, but not to a nucleic acid from
6 BCG.

1 26. A method of producing an attenuated Mycobacterium species, said
2 method comprising deleting from the genomic DNA of a virulent mycobacterium a first
3 nucleic acid that specifically hybridizes under stringent conditions with a second nucleic
4 acid or a complement of said second nucleic acid where said second nucleic acid or
5 complement of said second nucleic acid is selected from the group consisting of BCG Δ 1,
6 BCG Δ 2, and BCG Δ 3.

Figure 1

1 GAATTCCTGC GCACCTGAT CCTGTGCTG GTGCGATGA CTATCCAGA CTCTGCTAT 100
101 GAATGGAAGA CTTTCCGAC ACTGCGCTG TGTTCACCA CATGCGGAG GATGCTGCT 200
201 TGGCGCCG TCAATCTTC GATGAGCGG GATGAAAGT GCGCGCGCG GATGCTGCT 300
301 GAATACCC CCGTCCGAC GCTTTTCTG GTGTGCGAG GTGTGCGAG GTGTGCTG 400
401 GCGTCCGCG GTCCATCTG TGTGCTGAG GTGTGCTG GTGTGCTG GTGTGCTG 500
501 AATGCGATG GCGACACCA GCTCTCATG ATCCAGCGG GTATTCGCA GCGTCTGCT 600
601 GCGTCCGCA TGGAGACCC GGTCAAGTC AGCACCTTC ACATCAGTG GCGTCTGCT 700
701 GCGCGCGCA ACAGACCAT AGACAGCGG GCGCACCTC CAGTTCTAC GCGTCTGCT 800
801 GATCCAAAG CAGACATGA GAGAGCGG GCGCACCGG CCGCGCTGC GCGTCTGCT 900
901 CCGGATGAC GATGCTGCG GAGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCT 1000
1001 CCGTCAAGG GCTCATCGC CCGTATCGC GACACCTTC GCGTCTGCT GCGTCTGCT 1100
1101 AGACCTTTC GCGGCTGCG GCGACATCG TATTGCGCG GCGTCTGCT GCGTCTGCT 1200
1201 CACTACCGC GCAAGCTTC GTTCTATTG ATCCAGCTG GCGTCTGCT GCGTCTGCT 1300
1301 CCGACCGCA CAGGCTGAC CCGTGTGCG CAGAGTGA AGCGCTGCT GCGTCTGCT 1400
1401 GTACCGCG CCGCTGCGG ATCCAGTGA ACCGTTGCG TCCGATCTC GCGTCTGCT 1500
1501 TTTCCCGAC GTTACGCGA GGTTCAGAT CTGCGCGCG AGCGCTGCG GCGTCTGCT 1600
1601 GCGGCTGCG CCACTACCT GCGACACCA TCGAGTTGCG GTTGTGCG GCGTCTGCT 1700
1701 TCGGCTGCG GCGAGTTC GCGACATCG CCACTGCT GCGTCTGCG GCGTCTGCT 1800
1801 CCGGACCGA GTCCGCTAC CCGACTGCT GCGAGTTC GCGTCTGCG GCGTCTGCT 1900
1901 CCGTCTGCG CCGGCTGCG CCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 2000
2001 ATGCTGCTG CCACTGCTG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 2100
2101 CCGCTGCG CCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 2200
2201 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 2300
2301 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 2400
2401 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 2500
2501 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 2600
2601 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 2700
2701 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 2800
2801 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 2900
2901 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 3000
3001 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 3100
3101 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 3200
3201 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 3300
3301 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 3400
3401 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 3500
3501 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 3600
3601 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 3700
3701 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 3800
3801 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 3900
3901 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 4000
4001 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 4100
4101 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 4200
4201 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 4300
4301 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 4400
4401 GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGGCTGCG GCGTCTGCG GCGTCTGCT 4500

2/16

Figure 1 continued, page 2 of 4

4501 GTGCTGCGT TCGAGAGCC AGCCATTAG CAGAGCAGG AACTGACGA
 4601 AGGAGAGCA GAGAGGCGT TCGCTGCGA TCGCTTTCTA CCGCTGATC
 4701 GGTATCGAG CCGCGCGAG CCGATTCGA GGAATTCGA CGTCCATTA
 4801 GGGGCGTAG CGGCTGCGG GGTATTCGA GGTATTCGA AATATGAGC
 4901 CAGCGAGCC GGTTCAGCA TCGCTTCGC CTTTCTCGT GTTTATACG
 5001 CCGCGCGAT GAGCTTTCC GAGCTTCGA CCGATTCGA GCGACCGT
 5101 CCGCGCGAT GAGCTTTCC GAGCTTCGA CCGATTCGA GCGACCGT
 5201 CCGAGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 5301 GCGAGAGCC GATTCGAGT AGCGCGAGC AGCTTCAGG AGCTTCAGG
 5401 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 5501 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 5601 CCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 5701 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 5801 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 5901 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 6001 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 6101 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 6201 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 6301 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 6401 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 6501 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 6601 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 6701 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 6801 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 6901 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 7001 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 7101 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 7201 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 7301 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 7401 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 7501 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 7601 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 7701 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 7801 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 7901 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 8001 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 8101 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 8201 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 8301 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 8401 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 8501 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 8601 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 8701 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 8801 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 8901 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 9001 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 9101 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 9201 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 9301 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 9401 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 9501 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG
 9601 GCGCGAGCC AGCGCGAGC AGCTTCAGG AGCTTCAGG AGCTTCAGG

[illegible]

144901	CGTAAATCCGG	CGACGAGTCCG	GAACCCGCGTC	CACCCACCCGG	GGTGGAGATC	GTCTGACGTT	GTATGCGCGC	GAAGCAGCGC	CGGCGCGAAT	GTCCACAGTC	15000
150001	ACCGCGGCCA	CCCGATCCGAT	GAACCCGCGCG	CGACACGAGTT	CGTTCGACCC	GGCCCCCGGG	TGTTCTCCGG	TGGCGGTGAT	CGGTGAGATG	ATCTGCGCGC	15100
151001	CGGTTCAGGTC	GGCGAACCC	TGCGGAMCGA	GGGCGCCGAG	ACCGCAGACA	TATGTCGCGG	CAAAACGAGT	GGCGCGGATG	GGTACACCGC	CTCTCCGCGC	15200
152001	TTCCAAAGCA	TTACACCGGTT	CACCGCGGTC	CCCGAGGCGG	ACGATGTTTT	CTCCCGGGCG	GGGACMGTTCC	ACCCACGCTC	GGTGCATCGA	GGACGAGCTG	15300
153001	GGCATTCGCG	TCGTGCGCGAT	ACCGCCGAGC	CTTACGACCA	CGGTTCGGTA	CTCACCCGCG	GGGACMGCGG	TCTGCAATAT	GTTCCTACCG	CGTGGGTGCG	15400
154001	CGGATGTGGA	TCGTGCGCGAT	ACCGCCGAGC	CTTACGACCA	CGGTTCGGTA	CTCACCCGCG	GGGACMGCGG	TCTGCAATAT	GTTCCTACCG	CGTGGGTGCG	15500
155001	CGGATGTGGA	TCGTGCGCGAT	ACCGCCGAGC	CTTACGACCA	CGGTTCGGTA	CTCACCCGCG	GGGACMGCGG	TCTGCAATAT	GTTCCTACCG	CGTGGGTGCG	15600
156001	ACCAACGCGC	CCGAGTAGGG	TTCTATCGAT	CGGCTTGCTC	ACCTTGTAGC	AGCGCGCTTC	ACTGATGTTG	ATCACACCCA	CGCGGAGGAT	CGCTCGAGAG	15700
157001	CTATGAGCTAG	GGCGACAGCT	CGCGATCGAA	CGGTGCGCGC	GGGTGGCGTT	GGGTTCGATC	GGGTGGCGTT	GTGACGCGAC	CGGTTCCGAG	GGCTTCAGATG	15800
158001	GGCATTCAGAG	TCAGACGAGCT	CGGAGTCCGG	CGGCGACGCC	CGGTGCGCGC	GGGTTCGATC	GGGTGGCGTT	GTGACGCGAC	CGGTTCCGAG	GGCTTCAGATG	15900
159001	CGCAACGCTA	CCCCCGGCCC	GGTTCGCGAC	TTGTGGCGAT	CGGCCACGCC	CAGATTCGCG	CGGACMGCGT	GGCGACGCGT	CGACACCGGT	GTGCGATCACC	16000
160001	TGCTGGCGCG	CGCGCGACAA	ACCGCGTGTT	GGGTAGCGTG	ATCCGAGCGC	CGGATTCGCG	TTGCTCCAGC	CGCGATTCGT	GAACCGCGAC	CGCTGATGCG	16100
161001	GACCGCGCGC	CGCGCGGTGA	CGCGCGGTGA	CGGACGCGTG	CGGACGCGTG	TGCGATTCGCG	TTGCTCCAGC	CGCGATTCGT	GAACCGCGAC	CGCTGATGCG	16200
162001	TGACGCGCGT	TGCTTACGCTG	GGCGACCTGA	CGGACGCGTG	ATCCGAGCGC	TGCGATTCGCG	TTGCTCCAGC	CGCGATTCGT	GAACCGCGAC	CGCTGATGCG	16300
163001	CGCGACGCGC	CGCGTAGCGC	CTGTACTCGG	ACCAACGCGG	CGGACGCGTG	CGGATTCGCG	TTGCTCCAGC	CGCGATTCGT	GAACCGCGAC	CGCTGATGCG	16400
164001	ATGCTTCGCG	TGCTTCGAGCA	CAGCGCTGCT	ATCCGAGCGC	CGGACGCGTG	CGGATTCGCG	TTGCTCCAGC	CGCGATTCGT	GAACCGCGAC	CGCTGATGCG	16500
165001	TTGTGCGCGC	TCGCGCGGCG	TCGCGCGGCG	CGGACGCGTG	CGGACGCGTG	CGGATTCGCG	TTGCTCCAGC	CGCGATTCGT	GAACCGCGAC	CGCTGATGCG	16600
166001	AGGCGGTGCG	TAGTTCGCGA	TGCGCGGCGG	CGGACGCGTG	CGGACGCGTG	CGGATTCGCG	TTGCTCCAGC	CGCGATTCGT	GAACCGCGAC	CGCTGATGCG	16700
167001	ATCTTCGAGCA	TCTCCACCGG	CGAATAGACG	TGCAACCGCA	CGTTTCGGTT	CGGATTCGCG	TTGCTCCAGC	CGCGATTCGT	GAACCGCGAC	CGCTGATGCG	16800
168001	TTCCGAATCG	ATACCTTGGA	TCCGAGAGCG	TCCCAACCGC	TTGATTCGAT	ACGTTGAGTA	TTCTATATGG	TGCTACCAAT	GAGGAATTCG	GTACCCCGCG	16885

5/16

Figure 2

Page 1 of 4

1 GATCTCTGG ACTGCTGG TGCTGCTTG TGCAAGATT CACGAGTTC CACGAGTTC CACATCAAC CACTTCCGG 100
 101 GCGACGACG AACCTGATC GGAATATCC GGGTACGGA CGCGCGAT CGCGCGAT CACTGTGAGC AACAGTCCG 200
 201 GGTCTCTGG AGTTATACC CGTTATACT ATCTGATAC AGACGATAT GGTGAGTGT GTTGTGACCA CACTGTGAGC 300
 301 TGCTCTGG TGCTCTGG TTTTACGA AGCTTCCGA GCGTACTAT CAGGCTGG AGCTGAGAT TCGATTCGGT CAGATGAGC 400
 401 GAGCATCAT GGCACGAGC CTCTGCTGG GCGACGAT AGCGCTGG AGAACCATG CACGATGAT GGTGATTAGT GTGCTCGAGA 500
 501 TCTACTGCG TGACTTGG CCGCATGAG GCACTGAGC GCGGAGTCC GTTCTTCCG GCGACACAA CCGACTGCC AGCTGAGTCC 600
 601 CACTGTGATC GAGCGGTGA TACGCTGCA TACGCTGCA TACGCTGCA TACGCTGCA TACGCTGCA TACGCTGCA 700
 701 GTCTCTGAG TCTAGCTTC TCTAGCTTC TCTAGCTTC TCTAGCTTC TCTAGCTTC TCTAGCTTC TCTAGCTTC 800
 801 TCTGCTGAT ACTGCTTC TCTGCTGAT TCTGCTGAT TCTGCTGAT TCTGCTGAT TCTGCTGAT TCTGCTGAT 900
 901 GAGTGTGAT TGTGCTGAG TGTGCTGAG TGTGCTGAG TGTGCTGAG TGTGCTGAG TGTGCTGAG TGTGCTGAG 1000
 1001 AGCTGCTGG CTGCTGAG TGTGCTGAG TGTGCTGAG TGTGCTGAG TGTGCTGAG TGTGCTGAG TGTGCTGAG 1100
 1101 GTGCTGAGT CTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 1200
 1201 GCGAGCTGG GTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 1300
 1301 GCGCTGCTG TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 1400
 1401 CCGATGCTG TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 1500
 1501 ATGCTGCTG TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 1600
 1601 ATGCTGCTG TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 1700
 1701 GCGCTGCTG TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 1800
 1801 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 1900
 1901 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 2000
 2001 ATGCTGCTG TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 2100
 2101 GCGCTGCTG TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 2200
 2201 GCGCTGCTG TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 2300
 2301 ATGCTGCTG TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 2400
 2401 GCGCTGCTG TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 2500
 2501 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 2600
 2601 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 2700
 2701 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 2800
 2801 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 2900
 2901 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 3000
 3001 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 3100
 3101 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 3200
 3201 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 3300
 3301 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 3400
 3401 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 3500
 3501 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 3600
 3601 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 3700
 3701 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 3800
 3801 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 3900
 3901 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 4000
 4001 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 4100
 4101 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 4200
 4201 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 4300
 4301 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 4400
 4401 TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT TGTGCTGAGT 4500

Figure 2 continued, page 2 of 4

4501 TGTGACAGG TGTGTACTA AGGACGATC AGGACGAC CAGGCGAC 4600
 4601 TGAACAGCC CCGGCTCCG ACCGAGTCC AGGACGAC CAGGCGAC 4700
 4701 GTGTCTGTTT GCGACGCGC CTAGCAGAC GAGGTGTTC AGGACGAC 4800
 4801 GGAACGCGC TGGGCGGAG TGGGATGAG GAGGATGAG GAGGATGAG 4900
 4901 TGAACGACAT GAGGCGAGC GAGGATGAG GAGGATGAG GAGGATGAG 5000
 5001 CAGGATGAG AGGATGAGC GAGGATGAG GAGGATGAG GAGGATGAG 5100
 5101 AGGATGAGC AGGATGAGC GAGGATGAG GAGGATGAG GAGGATGAG 5200
 5201 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 5300
 5301 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 5400
 5401 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 5500
 5501 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 5600
 5601 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 5700
 5701 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 5800
 5801 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 5900
 6001 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 6100
 6101 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 6200
 6201 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 6300
 6301 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 6400
 6501 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 6600
 6601 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 6700
 6701 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 6800
 6801 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 6900
 6901 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 7000
 7101 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 7200
 7201 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 7300
 7401 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 7500
 7601 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 7700
 7701 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 7800
 7901 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 8000
 8101 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 8200
 8201 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 8300
 8401 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 8500
 8601 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 8700
 8701 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 8800
 8801 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 8900
 8901 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 9000
 9101 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 9200
 9201 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 9300
 9401 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 9500
 9501 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 9600
 9601 GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC GGTCTGAGC 9700

9701	GCACGCGTC	CCGTAACCGT	ATCCGCGGTG	ATCCGCGGTG	GTAAATGAGTT	CAGCGACAGG	AGAAATAAA	TATGCTAGG	CGAAATACAT	AGCTACAGT	9800
9801	CCCTGTGCAC	CCCATGAAG	ACCTCAATG	ACAGAGAGA	CGCGCGAAAT	GCCTCAGCGG	CGACTGCCG	ACTTACGCC	CACAGBGG	TACCCAGCGG	9900
9901	CCAAACAGCG	CATGACATG	CGAAGGAGG	ACGCGCCCTG	ATGACCGCGA	AGCGGTTTGA	GGCGGTGCAC	ACCTGCATAT	TTCGACAAAT	CGTCAAGCTG	1000
0001	AAAGATTTCG	AGGTCCTGGA	ACGTTTGACC	GGTAATCTTT	GGTTTGCAGC	AAAGATTTCG	CTCTCCAGAC	ACCTGCAGAT	TTCGACAAAT	TTGAGTTCCA	1010
0101	GAAGGACAGA	GAAGGAGATC	CGCGTTTCAC	CGCGTTTCAC	CGCTCTGCAC	ACCGCGAGAG	CGACGCGTGG	AGCATGTGCC	ATCATTCBAG	ACCGGCTAC	1020
0201	CCCCACAGAA	GAAGCGGTCC	TGACAAATAT	CGGTTTCATG	CGCTCTGCAC	AGCGCGAGAG	CGACGCGTGG	ATCTTCTCGA	CCCTGTGCTC	CATCAAGCGA	1030
0301	ATGACAGATG	CTTTCAGACT	GTCCGATAC	GAAGCTTACC	TGCAAGCGAA	AGCGCGAGAG	CTACAGCTCG	ACTAACCGCG	TTCAGACCGG	CTCAAGCGCA	1040
0401	AGCATCTGTC	GGTAATAGTG	GAGTCTTTCG	TGTTCTTACT	CGCTCTGCAC	CTCCCGCATG	ACTGTGTGTC	GGCGGTGAAG	CTCAACAGCA	CGCGCGATCT	1050
0501	GATCCGCGAT	ATCATCTGAG	ATGAGCGATC	ATGAGCGATC	TACATCTGCAT	ACAAATGTGA	ACGAGGTTTG	GGCAGCTTGA	CGACGCGCA	GGCGCGCGAG	1060
0601	AGCATCTGAT	ACATCTGCGA	CGCTCTGCAC	AGCTCTTACG	CGAAGAGAT	CGACTATGCG	CAGACTGTGG	ACGACGAGTT	GGCTGTGAGC	GAGCAGCTTT	1070
0701	TGCCCTTACAT	GGTTTACAAAC	CCCAACAGAG	CCCAACAGAG	CTCGGATAC	GAAGCTCAT	TGCAATGTGA	CACCTGCCAG	GTGACAAACG	CGCTGTGCGT	1080
0801	AGCTCTGCAC	CCCGGTGCGA	GGGAGAGCA	CGACTTTTTC	CTCGGATAC	GAAGCTCAT	CTGTAATGGC	ACCAACCAAC	CCACACCGA	CACGACCTGG	1090
0901	GACTCTTAC	CCCGGTGCGA	GGGAGAGCA	CGACTTTTTC	CTCGGATAC	GGCCCGATCG	ATCTGTCTAG	TGTAGTCTGG	TTCAGCATCG	TGCTGTACGAT	1100
1001	CCGATGCGCC	TATGTTTGTG	GTGACTCTAG	ATATCTGCGC	ATATCAATCT	TCGCGGATCG	GGCGCGGTGG	CTGCAACTA	CTCCCGAGG	ATTTTCTGAG	1110
1101	TGCGGATCAA	GATCTCTATG	CTGTCTACGG	CTGTCTCTAG	CTCTGTCTGT	TGCGGATCG	GGCGCGGTGG	CTGCAACTA	CTCCCGAGG	ATTTTCTGAG	1120
1201	GGCGGATCAC	GGCGGATCAC	CCGATCTTCA	ATATCTGCGC	ATATCAATCT	TGCGGATCG	GGCGCGGTGG	CTGCAACTA	CTCCCGAGG	ATTTTCTGAG	1130
1301	ATTTACATCG	CCGAGACCGG	CGACAGTTTC	CTCAGCGCGG	CCACATGCTC	AGCTCATGAT	CAGCTCTGCC	GGCGCGGTGG	CTGCAACTA	CTCCCGAGG	1140
1401	AGTCCCGGAT	ACCGCGCGGT	GGTACGAGG	CGGTCGAGG	CAGGCTCTAC	CAGACACCGG	GGCGCGGTGG	CTGCAACTA	CTCCCGAGG	ATTTTCTGAG	1150
1501	AGTCCCGGAT	ATTCGCGACG	ATTCGCGACG	ATTCGCGACG	ATTCGCGACG	AGACAGCGCG	GGTACGAGG	CGGTCGAGG	CGCTCTGATG	CTTCGATTTG	1160
1601	AGACAGAGCG	GAACAGAGCT	ATTCGCGACG	ATTCGCGACG	ATTCGCGACG	GGTACGAGG	CGGTCGAGG	CGGTCGAGG	CGCTCTGATG	CTTCGATTTG	1170
1701	TCACACCGCG	GGATTTGCTG	CTCCGAGCGG	CTCCGAGCGG	CTCCGAGCGG	GGTACGAGG	CGGTCGAGG	CGGTCGAGG	CGCTCTGATG	CTTCGATTTG	1180
1801	CAGAGATCTG	CTCTAGAGAA	CTCTAGAGAA	CTCTAGAGAA	CTCTAGAGAA	GGTACGAGG	CGGTCGAGG	CGGTCGAGG	CGCTCTGATG	CTTCGATTTG	1190
1901	AGCTGTGCTG	TCCTAGAGAA	CTCTAGAGAA	CTCTAGAGAA	CTCTAGAGAA	GGTACGAGG	CGGTCGAGG	CGGTCGAGG	CGCTCTGATG	CTTCGATTTG	1200
2001	AGCTGTGCTG	TCCTAGAGAA	CTCTAGAGAA	CTCTAGAGAA	CTCTAGAGAA	GGTACGAGG	CGGTCGAGG	CGGTCGAGG	CGCTCTGATG	CTTCGATTTG	1210
2101	TCGCGGATCA	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGTACGAGG	CGGTCGAGG	CGGTCGAGG	CGCTCTGATG	CTTCGATTTG	1220
2201	TCGCGGATCA	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGTACGAGG	CGGTCGAGG	CGGTCGAGG	CGCTCTGATG	CTTCGATTTG	1230
2301	TCGCGGATCA	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGTACGAGG	CGGTCGAGG	CGGTCGAGG	CGCTCTGATG	CTTCGATTTG	1240
2401	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGTACGAGG	CGGTCGAGG	CGGTCGAGG	CGCTCTGATG	CTTCGATTTG	1250
2501	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGTACGAGG	CGGTCGAGG	CGGTCGAGG	CGCTCTGATG	CTTCGATTTG	1260
2601	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGTACGAGG	CGGTCGAGG	CGGTCGAGG	CGCTCTGATG	CTTCGATTTG	1270
2701	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGTACGAGG	CGGTCGAGG	CGGTCGAGG	CGCTCTGATG	CTTCGATTTG	1280
2801	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGCTGTGCTG	GGTACGAGG	CGGTCGAGG	CGGTCGAGG	CGCTCTGATG	CTTCGATTTG	1290
2901	GGCTGTGCTG	GGCTGTGCTG</									

8/16

Figure 2 continued, page 4 of 4

```
14901 TCATGTATCT TCTAGTATC CTATGCCCG GCGCATGTG CTCTCTGCC TTGCGCGCG ACAGCGCAT CACCGTCAA GCGTTGGCT CAGTCAGGC 15000
15001 TGGGAGGCA CCGCATGAG GATCAGCAG CCGCGCGAC GCGCGTGAT CCGATCGCG ATCTGACCG CTCTGATTCG GCGCGCGCA CAGAGGCTT 15100
15101 CCGGACCGC CAGGTTCGA CCAATGCTT CAGAGCACT TCTCTCTAC GAGCTGTAG TTTTCCGCG CCCACTCCG GCTTTGTCTG CAGGTTTCA 15200
15201 CCGTCAGGCA CCGCTCATG CATTGCGAC GCGGAATTC
      | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100
```

Figure 3

1	GAATTCACCTT	AGCTATACAC	AGTTCTACGA	CGTGTGGCG	CGACTTCTTG	TCAGTGGCG	ACGTTATGAT	TCGAACATGT	TAGCGAATAG	CCGCGAGGAG	100
101	CTTGTGGAGG	TCCTTTGATG	CGTGGATGCG	GAGCTGGACC	CGTTGTGAGC	GGTGTCTTTT	GAGGTGTGTG	CGACCCCGGA	ACGCTGGCG	TCCTTGGAAC	200
201	CTTGTGGCGG	CGGCTTACCG	CGGTCGGGCA	CGAGTGGGCA	CAGCTTGATC	GACCACTCGT	ACACCCCAAG	CAGCGAGGAA	GAGCTGGCG	GCAGCTGTGG	300
301	GTCTCGCGATG	GCCAACTCAG	TACGCAATAC	CGAGTGGGCA	CGCCCTCTAC	GCATTCGCGA	CGCCGCGGAT	GTCGCGAGAC	GTCCGAGGAC	TCACCGCGGA	400
401	ACCGCTAGCG	CCACAGTTTG	ACCGGCACCG	CCACCGCCCA	ACGGCAGGGC	CTGANTCGCG	MAGCGGCACA	TCAAAGTGAT	TCCGCGCCCTT	TTTTCGGCCA	500
501	ACTTGCGCCG	CGCGTGTGAT	GTGTGCMAAC	CCCGTCAGGC	CGCGAGGGCC	GACCTTGGCG	CAAAACCGCT	AAATATCGTC	CCGACGAGCT	CGACCCGTAC	600
601	CTACCGAGCG	TCATGAGTGC	CGTACACCCC	GACGGCGACC	TCACCGAGAC	GAACCGCGCC	CGATACCCCT	GAGCMAACAG	CGATACCCCT	GAATACGACG	700
701	CGATGTACCG	TACGTAGCCG	CGCATGACCG	CGCAAGGCGC	GGCCACTCTT	GAAGCGGTGC	TACGCAAACT	GGCGCCCGCC	CGCGGAGGCA	ACCCCGAGCA	800
801	CTCAACCCCG	GTATATGACA	CCACCCCGCA	TGCGCGCGCC	ATCGAACCGG	ACACCCCGAG	CGAAGCCCAA	CGCAACCCAG	ACCGGCTGCT	GGCGCGCGTG	900
901	CGCGCGCTGA	TCCGCTCCCG	GGAACTGGGC	CAACACCAAG	GTGTATCCGT	ATCGAGTGTG	GTGACGAGCA	CCCTGACCCA	CGCTGCAAAAC	GGCGCCCGCA	1000
1001	AGGCGTTTAC	CGCGCGCGGC	ACCGCTGGTC	CGATGTGGCA	TGTTATCCGC	ATGACAGACC	AGCGCCAGCA	CCCTATCCCG	CGTAAGCGGA	GGTACCCCCA	1100
1101	CGACCTCTTC	GACCAACCGCA	CACCCCTGGC	CGTGTATGAC	TCGTTATCCG	TAGCCTTCCC	AGCGCCAGGG	ATCATGTGTT	TGCGCAAGGA	CGCGGCTTGC	1200
1201	ACCAAACCCG	CGCTGTAGCG	ACCGGCTTAC	CAACGCGAAG	CCACACAGTT	CACCGGCTGG	ACCGAGCAGG	GAGCGACCGA	CATCACCGAC	CTCACCCCTG	1300
1301	CTCGCGACCC	CGACAAAGCA	CTCGCGGAAA	AGGCTGTGAC	CACCCGCAAA	ATCAAGCAAT	CGCAACAGCA	ATGCGTACCA	CGACCCCAAC	CTGACCAACG	1400
1401	CCACCTGTGAC	ACCTGTGAGA	TACACTACAC	CTGTGCGTGC	TGCTGTCTTAC	CGCTGTAACT	CAGAGAGCGA	CTCGCGGMA	CAGCTGGCG	TGCGCCCGCA	1500
1501	ACCGCGGAGC	TGCTGTAGCA	TGTGTGGCGA	CGCAAAATGG	TCGCTGTCTAC	AGTAACTTCA	CACAGAGCTC	AGCGCATCAA	CGCGCAAGCG	CGCGCCCGCC	1600
1601	TACAGAGCAG	TGTTTGCCGA	CATCAACCGC	GGCAAGATCG	CGCGCTGTGT	GGCTTGGAG	CTGGACCGGC	TCCANTCGCG	TCCANTCGAG	CTCGAAGCCT	1700
1701	TCATGTCTATT	AGCGCAAGAG	AGCGCGCTGG	CCCTGTGCAC	CGTCCGCGCG	GAGCTTGACC	TGGCGACACC	CGAGCGCCGG	CTATGTGCCC	CGTGAAGGCG	1800
1801	GTCCGTTGCC	CGTCAGACAA	CGAGGACAAA	GAGCGGCGCG	CAGCGCCGCG	CGCGCCCGCA	GAAGACTGMA	CGCGGCCAC	CGCACTGGTC	GAAGACTGTC	1900
1901	CGCGCCCGCA	CGCGCTCCGA	CGCGCCCGCC	GGACAGCGCC	GGACAGCGCC	CGCTGTCAAA	GAAAGCTTAC	CGCACTATCT	CGCACTATCT	TCCCTGGCGG	2000
2001	ACGTTGTGCG	CGAGTGGAAC	GACGCGCGGG	GTTTACCAAT	CACCGCCGCG	CGGTGGAGGA	CTTACACGCT	GTGCAAAATC	TTGCGCAAAC	CCCGCAACGC	2100
2101	CGGACTAGCG	GCATATATGG	GTGCGCGCTA	GGCGCGCGTG	GACCGCGAGG	CGATTTGTGG	CAMGGCCGAG	TGTTGTGCGC	TGTTGTGAGGA	GGCAACGTTT	2200
2201	TGGCGCGGCC	GGACCGCGCC	GGCGCGCGCC	GGCGCGCGCA	CGCGCGCGGA	AAAGTGTGGC	CGCGACCTGT	GGCACTGGCT	GGCAAGCTGC	GGCAAAATCG	2300
2301	CGACCGCGCG	TACCGCGCAG	ACGCGCGCGC	ACGCGCGCGC	CGTCTACGTG	TGCGAAGGCT	GGCAACGGGT	GGCATCTCTG	GGCAACCAICA	TGCAACCGAT	2400
2401	ATCGTGGCGG	AGCGGCTGCG	CATGCGCGAC	CGCTTGTACT	CGCTTGTACT	TGTTTGGCGG	GGAGATTAC	GACCGCGCG	AGCGCGAAC	CATCGCGCTG	2500
2501	CCCTCTACCG	GAGCTGGACA	GGCTCGCGCT	CGACCGCGCC	CGACCGCGCC	GAAAGGCTAC	TGACCGCGCG	CTCAGTTAAG	ATCAGCAACG	ACATCTGTCA	2600
2601	CGCGCAGATA	AGCGCAATTC	AGCGCCCGCA	AGCGCGCGCA	AGACGCTCTC	GAGTGTTTGA	CGGATTAAG	TTGGGAAAC	CGGAGTGGC	CGGAGTATA	2700
2701	CGCGCAGATA	AGCGCAATTC	AGCGCCCGCA	AGCGCGCGCA	AGCTCTCTGA	AGTGTGTTTC	CAGCGCGTGG	CGAAGAGCGG	CGGATATATC	ATATCGCAAC	2800
2801	CGCGCAGATA	AGCGCAATTC	AGCGCCCGCA	AGCGCGCGCA	AGCTCTCTGA	AGTGTGTTTC	CAGCGCGTGG	CGAAGAGCGG	CGGATATATC	ATATCGCAAC	2900
2901	CGCGCAGATA	AGCGCAATTC	AGCGCCCGCA	AGCGCGCGCA	AGCTCTCTGA	AGTGTGTTTC	CAGCGCGTGG	CGAAGAGCGG	CGGATATATC	ATATCGCAAC	3000
3001	CGCGCAGATA	AGCGCAATTC	AGCGCCCGCA	AGCGCGCGCA	AGCTCTCTGA	AGTGTGTTTC	CAGCGCGTGG	CGAAGAGCGG	CGGATATATC	ATATCGCAAC	3100
3101	CGCGCAGATA	AGCGCAATTC	AGCGCCCGCA	AGCGCGCGCA	AGCTCTCTGA	AGTGTGTTTC	CAGCGCGTGG	CGAAGAGCGG	CGGATATATC	ATATCGCAAC	3200
3201	CGCGCAGATA	AGCGCAATTC	AGCGCCCGCA	AGCGCGCGCA	AGCTCTCTGA	AGTGTGTTTC	CAGCGCGTGG	CGAAGAGCGG	CGGATATATC	ATATCGCAAC	3300
3301	CGCGCAGATA										

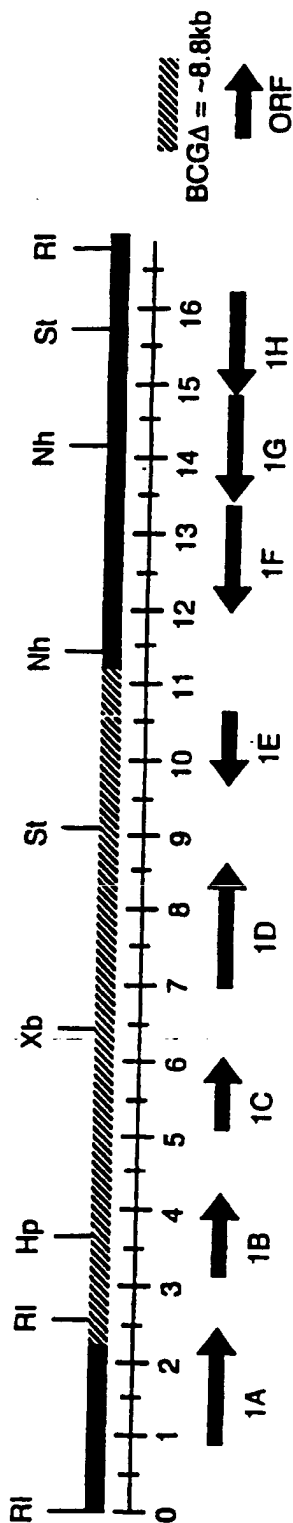
[illegible]

11/16

Figure 3 continued, page 3 of 3

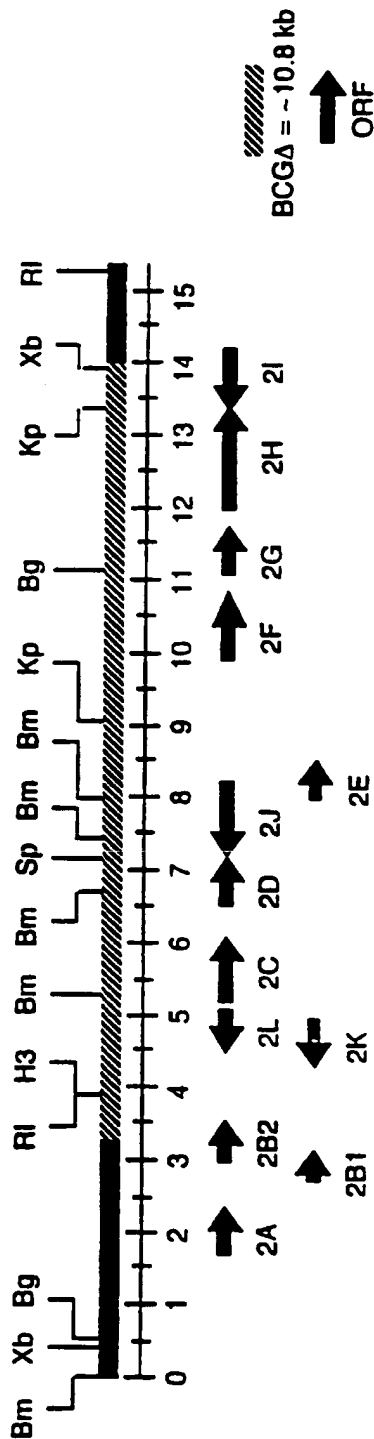
9701	CTCTCTTTC	AGTGTGCG	TGAGAGGCT	TTGGTTGAT	GCGACCAAG	CCGCAATTGG	TGATGATGC	AATGAGGCA	CGACCTTGC	GTTTGGGCTC	9800
9801	CTTGGCTTC	GCGGTTTCA	GCTGCGGAT	CCGCGACCA	GCGCGGAT	TTGTTTCTC	GCGAACCGC	CGATTCTTT	GTCGTGCTG	TGGGTTGGCT	9900
9901	TGTGATTTG	TTTGAGCGC	CGGTAGTGC	CGCGGAGAT	CGCGGAGCT	CGCGGATAG	CAGTGTCTG	TTTGTGCGA	TGACGCTCTG	ATTCTGGTTT	10000
10001	GTAGCCATG	GCGGCTCAAC	CAGATATCG	AGCTTGGCG	CGATGATCT	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GCGCGCTCG	CTCGCGGATC	10100
10101	CGCGGCGAG	GCGGCGCAT	TTTCTCTCG	AGATAGCTT	CGCGGATAG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	AGCGCGGAT	GTCCTGTGCT	10200
10201	CGCGGCGAG	AGCGGATTC	CTGCTCTTC	GCTGCGCTT	CGCGGCGAG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	10300
10301	CGCGGCGAG	TGCGGATTC	CGGTTGATG	GCGCTTCTT	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	10400
10401	GTCGTGCTG	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	10500
10501	GCGGATTC	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	10600
10601	TGAGGATTC	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	10700
10701	AGCTGCTTC	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	10800
10801	AGCTGCTTC	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	10900
10901	AGCTGCTTC	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	11000
11001	GTCGTGCTG	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	11100
11101	GTCGTGCTG	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	11200
11201	GTCGTGCTG	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	11300
11301	GTCGTGCTG	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	11400
11401	GTCGTGCTG	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	11500
11501	GTCGTGCTG	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	11600
11601	GTCGTGCTG	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	11700
11701	GTCGTGCTG	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	11800
11801	GTCGTGCTG	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	11900
11901	GTCGTGCTG	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	12000
12001	GTCGTGCTG	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	12100
12101	GTCGTGCTG	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	12200
12201	GTCGTGCTG	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	12300
12301	GTCGTGCTG	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	12400
12401	TGCGGATTC	GCGGATTC	GCGGATTC	GCGGATTC	GTCGTGCTG	CGCGGATAG	CTGCGCTGC	TTGCGCTTC	GTCGTGCTG	GTCCTGTGCT	12412

12/16

Region 1 (16.9 kb)

ORF	<i>M. tuberculosis</i> Codon Usage	ORF Size (base pairs)	Start - Stop (base pairs)	Possible Ribosome Binding Sites	Encoded Protein (max. -kDa)	Homologies to Predicted Encoded Protein	P value	Homologue Accession #
1A	yes	1542	889 - 2433	AGGA (10)	57			
1B	yes	1071	3130 - 4203	GGA (4)	36	<i>M. leprae</i> aceA	1.4e-14	Z46257
1C	yes	969	3139 - 4203	GGA (9)	36	BCG uraA	3.0e-13	U01072
1D	yes	1657	5075 - 6046	GAGG (5)	34	<i>M. tuberculosis</i> esa16	2.3e-43	X78562
1E	yes	954	6954 - 8612	none	59			
1F	yes	1380	10619 - 9663	GGA (5)	34			
1G	yes	1386	14823 - 13438	AGGA (9)	48			
1H	yes	1368	14643 - 13438	GGG (11)	50			
			14541 - 13438	AGGAGA (10)				
			16190 - 14820	GAA (5)	46	<i>B. subtilis</i> subtilisin Serine proteases	3.6e-16 4.8e-14	L29506 A08331

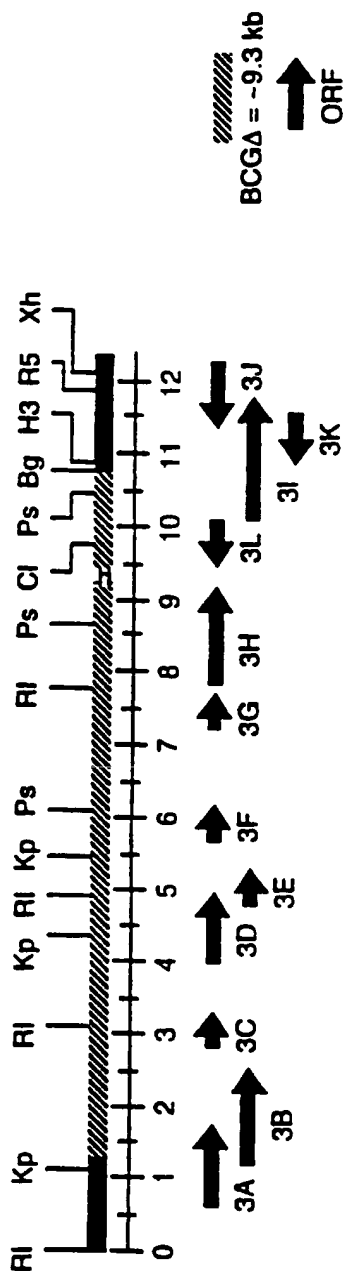
Figure 4

Region 2 (15.3 kb)

ORF	<i>M. tuberculosis</i> Codon Usage	ORF Size (base pairs)	Start - Stop (base pairs)	Possible Ribosomal Binding Sites	Encoded Protein (max. - kDa)	Homologies to Predicted Encoded Protein	P value	Homologue Accession #
2A	yes	558	1829 - 2386	AGGGAG (7)	25			
2B1	yes	437	2862 - 3298	AGAA (4)	16			
2B2	yes	588	3003 - 3590	none	34			
2C	yes	948	5187 - 6134 5376 - 6134	AG7 (8) GGA (8)	34	<i>E. coli</i> tclA lysR family	9.9e-47 <1e-5	P24194
2D	yes	657	6561 - 7217	none	22	<i>Mleprae</i> cosmid B1620 ORF Cutinases	1.5e-7 ~4e-5	U00015 A00975 U03393
2E	yes	522	8036 - 8560	none	19			
2F		966	9941 - 10909	AGGA (11)	37	<i>S. typhimurium</i> RNDPR proUVWX	9.9e-146 2.7e-36	X73226 X17445
2G	yes	668	11118 - 11783	AAGA (6)	24	<i>M. tuberculosis</i> mpl64	6.7e-141	A30545
2H	yes	1443	11965 - 13407	AG (10)	51	<i>E. coli</i> gabP permease <i>S. typhimurium</i> asp permease <i>T. harzianum</i> indol gene retroviral receptor	3.1e-11 1.4e-08 4.4e-11 2.5e-09	X65104 U04851 Z22594 X59155
2I	yes	848	14221 - 13376	GGAAGA (6)	31			
2J	yes	1050	8259 - 7211 7939 - 7211 7931 - 7211	GAG (10) GGAA (8) GGA (9)	35			
2K	yes	686	4992 - 4327	none	25			
2L	yes	597	5117 - 4521	AG (10)	21			

Figure 5

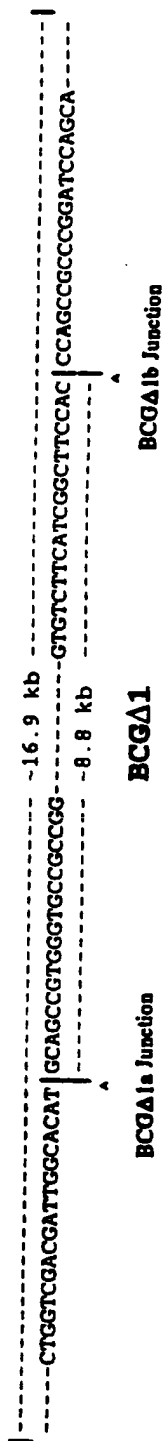
14/16

Region 3 (12.4 kb)

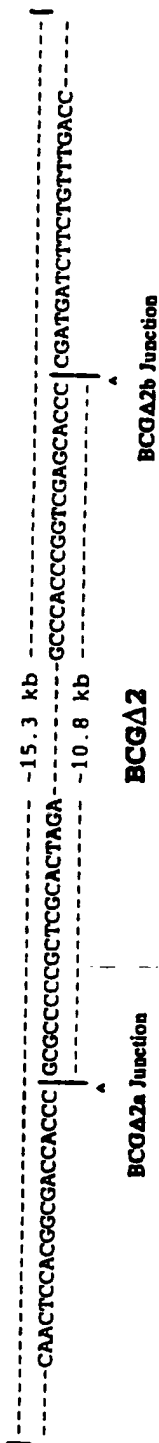
ORF	<i>M. tuberculosis</i> Codon Usage	ORF Size (base pairs)	Start - Stop (base pairs)	Possible Ribosomal Binding Sites	Encoded Protein (max. -kDa)	Homologies to Predicted Encoded Protein	P value	Homologue Accession #
3A	yes	1143	613 - 1755	none	42	MTB mce sau3A <i>M. leprae</i> cosmid L247	2.9e-64 5.1e-13	X70901 U00021
3B	yes	1347	1214 - 2560		49	Actinophage R4 altP gene <i>B. subtilis</i> site spec. recomb. recombinases / Invertases	3.0e-05 7.8e-4 8.2e-4	D90361 M29040 K00678 X01805, X07724
3C	yes	513	2820 - 3332	GGA (6)	19	<i>S. coelicolor</i> phage phi-C31 early region	4.2e-26	X76288
3D	no	924	4007 - 4930 4070 - 4930	none GGA (7)	34	<i>S. coelicolor</i> phage phi-C31 pglY pglZ genes	3.2e-11	X76288 L37531
3E	no	543	4795 - 5337 4915 - 5337	none GGAA (5)	21			
3F	yes	576	5639 - 6214	GA (9)	20			
3G	yes	510	7253 - 7762 7285 - 7762	GAAGG (8)	19			
3H	yes	1330	7868 - 9197	GA (8)	47			
3I	yes	1665	10146 - 11810 10164 - 11810	GGAAG (6)	58	<i>M. leprae</i> B1170	6.2e-69	U00010
3J	yes	918	12319 - 11402	GAA (11)	30	<i>M. leprae</i> bioDAYB <i>C. glutamicum</i> bioD <i>B. sphearicus</i> bioDAYB	6.9e-53 2.6e-08 1.4e-05	U00010 D14083 M29292
3K	yes	702	11594 - 10893	AGG (4)	25	<i>M. leprae</i> cosmid B1170	1.0e-81	U00010
3L	yes	680	10147 - 77 9488	AG (10)	25			

Figure 6

Region 1



Region 2



Region 3

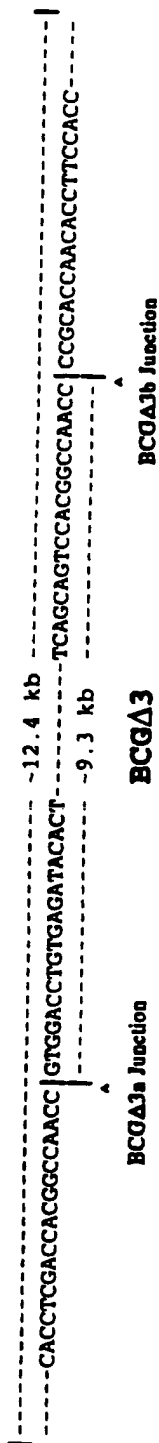
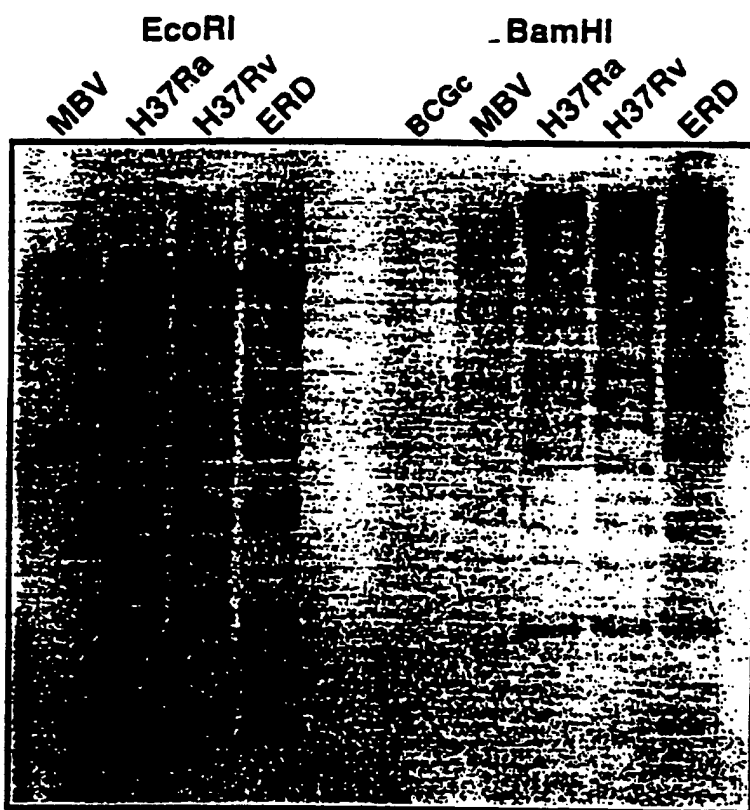


Figure 7

16/16

**Figure 8**

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/01938

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : Please See Extra Sheet.

US CL : Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 7.1, 91.1, 91.2, 240.2, 252.3; 530/300, 350, 387.1, 388.1; 536/22.1, 23.1, 24.5, 24.32, 24.33

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ---- Y	Infection and Immunity, Volume 61, No. 5, issued May 1993, H. Li et al, "Evidence for absence of the MPB64 gene in some substrains of Mycobacterium bovis BCG", pages 1730-1734, see entire document.	1-10, 16, 17, 24, 25 ----- 18-23
X	JP, 1-247094 (AJINOMOTO ET AL) 02 October 1989, see entire document.	1-7
X	Infection and Immunity, Volume 59, No. 10, issued October 1991, C. Parra et al, "Isolation, characterization and molecular cloning of a specific mycobacterium tuberculosis antigen gene: identification of a species-specific sequence", pages 3411-3417, see entire document.	1-17

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

17 APRIL 1996

Date of mailing of the international search report

29 MAY 1996

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Authorized officer

JEFFREY FREDMAN

Facsimile No. (703) 305-3230

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/01938

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Abstracts of the 1994 IDSA Annual Meeting, Clin. Infect. Dis., Volume 19, issued October 1994, R. Frothingham et al, "Sequence based strain differentiation in the Mycobacterium tuberculosis complex, including rapid identification of M. bovis BCG", page 565, see abstract 10.	1-25
X	R. GHERNA et al, "AMERICAN TYPE CULTURE COLLECTION: CATALOGUE OF BACTERIA AND PHAGES", Eighteenth edition, published 1992, pages 202 and 211, see entire document.	11-15
X	Infection and Immunity, Volume 62, No. 4, issued April 1994, L. Pascopella et al, "Use of in vivo complementation in Mycobacterium tuberculosis to identify a genomic fragment associated with virulence", pages 1313-1319, see entire document.	1-7, 16-25 ----
Y		26
Y	Science, Volume 261, issued 10 September 1993, S. Arruda et al, "Cloning of an M. Tuberculosis DNA fragment associated with entry and survival inside cells", pages 1454-1457, see entire document.	1-23
X	US,A,5,171,839 (PATARROYO) 15 December 1992, columns 5-10.	1-10 ----
Y		16-23
Y	Nature, Volume 256, issued 07 August 1975, C. Kohler et al, "Continuous cultures of fused cells secreting antibody of predefined specificity", pages 495-497, see entire document.	10
Y	US,A, 4,683,202 (MULLIS) 28 July 1987, see entire document.	16-22, 24, 25
Y	Genomics, Volume 4, issued 1989, D. Wu et al, "The ligation amplification reaction (LAR) amplification of specific DNA sequences using sequential rounds of template directed ligation", pages 560-569, see figure 2.	16-22, 24, 25
Y	US,A, 4,410,660 (STRAUS) 18 October 1983, columns 14 and 15.	23
Y	Gene, Volume 131, issued 1993, A. Kinger et al, "Identification and cloning of genes differentially expressed in the virulent strain of mycobacterium tuberculosis", pages 113-117, see page 114, column 2.	1-26
X,P	WO,A2,95/17511 (JACOBS ET AL) 29 June 1995, see entire document.	1-26

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US96/01938

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,E	J. Bacteriol., Volume 178, No. 5, issued March 1996, G. Mahairas et al, "Molecular analysis of genetic differences between mycobacterium bovis BCG and virulent M. bovis", pages 1274-1282, see entire document.	1-26
Y, P	Microbiology, Volume 141, issued 1995, J. Rodriguez et al, "Species-specific identification of mycobacterium bovis by PCR", pages 2131-2138, see entire document.	1-7, 16-22, 24, 25
X --- Y	Hybridoma, Volume 13, No. 1, issued 1994, A. Arya et al, "Production and characterization of new murine monoclonal antibodies reactive to mycobacterium tuberculosis", pages 21-30, see page 27, table 1.	8-10 --- 16-23

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/01938

A. CLASSIFICATION OF SUBJECT MATTER:

IPC (6):

C12Q 1/68; G01N 33/53; C12P 19/34; C12N 5/10, 1/21; C07K 5/00, 14/00, 16/00; C07H 21/02, 21/04

A. CLASSIFICATION OF SUBJECT MATTER:

US CL :

435/6, 7.1, 91.1, 91.2, 240.2, 252.3; 530/300, 350, 387.1, 388.1; 536/22.1, 23.1, 24.3, 24.32, 24.33

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, MEDLINE, BIOSIS, CAPLUS, WPIDS

search terms: mycobacter?, tubercul?, bovis?, BCG, calmette, guerin, DNA, RNA, oligo, nucleic, oligonucleotide, hybrid?, probe, primer, amplif?, PCR, polymerase chain, ligase chain, LCR, attenuat?, immunoassay, antibod?, monoclon?, polyclon?, protein, peptide, antigen, virulenc?, infect?